Biosynthetic Production and Structural Studies of Glycolipids and Large Fragments of G-protein Coupled Receptors

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Sowmini Kumaran

aus

Indien

Promotionskomitee

Prof. Dr. Oliver Zerbe (Vorsitz und Leitung der Dissertation)

Prof. Dr. John Robinson

Zürich, 2010

Table of Contents

Abbreviation

AM	adrenomedullin
AMP	Antimicrobial peptide
AR	adrenergic receptor
CARA	computer aided resonance assignment
CCR2	chemokine C-C motif receptor 2
CCR5	chemokine C-C motif receptor 5
CGRP	calcitonin generelated peptide
CD	circular dichroism
CD14	Cluster of differentiation 14
cDNA	**c**omplementary deoxyriobonucleic acid
CHAPS	3-(3-Cholamidopropyl)dimethylammonio)-1-Propanesulfonic Acid
CMC	critical micelle concentration
CNS	central nervous system
CV	column volume
CX3C	chemokine C-X3-C motif receptor
Dab	Diamino butyric acid
DCM	dichloromethane
DDM	dodecyl-β-d-maltoside
DHPC	1,2-Diheptanoyl-sn-Glycero-3-Phosphocholine
DIEA	diisopropylethylamine
DMF	dimethyl formamide
DPC	dodecylphosphocholine
Doxyl	(4,4-dimethyl-3-oxazolidine-N-oxyl)
DTT	1,4-dithiothreitol
ECL	extracellular loop
E.coli	*Escherichia coli*
EDTA	ethylenediamine tetraacetic acid
ELISA	enzyme linked immunostaining assay

ESI-MS	electrospray ionization mass spectroscopy
Fmoc	9-fluorenylmethoxycarbonyl
FPLC	fast protein liquid chromatography
GABA	gamma amino butyric acid
GdmHCl	guanidinium hydrochloride
GDP	guanosine diphosphate
GPCR	G-protein coupled receptor
GST	glutathione S transferase
HATU	2-(1H-7-Azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HBTU	2-(1H-benzotriazol-1-yL)-1,1,3,3-tetramethyluronium-hexafluorophosphat)
HEPES	2-(4-(2-hydroxyethyl)- 1-piperazinyl)-ethanesulfonic acid
HOBT	N-hydroxybenzotriazole
His-tag	Histidine tag
HPLC	high performance liquid chromatography
HSQC	heteronuclear single-quantum correlation
ICL	intracellular loop
IPTG	isopropyl β-D-1-thiogaloctopyranoside
K_D	dissociation constant
kDa	kilo Daltons
Kdo	3-deoxy-D-manno-oct-2-ulosonate
KSI	ketosteroidisomerase
LB	Luria bertani medium
LC-MS	liquid chromatography mass spectroscopy
LDAO	Lauryldimethylamine-oxide
LMPG	1-Myristoyl-2-Hydroxy-sn-Glycero-3-[Phospho- rac-(1-glycerol)]
LPPG	1-Palmitoyl-2-Hydroxy-sn-Glycero-3-[Phospho- rac-(1-glycerol)]
LPS	Lipopolysaccharide
MALDI-TOF	matrix-assisted Laser desorption ionization-time of flight
MBP	maltose binding protein

MD	Molecular dynamics
mCD14	Membrane bound CD14
MCP-1	monocyte chemoattractant protein 1
MES	4-Morpholine ethane sulfonic acid
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser enhancement spectroscopy
NPY	neuropeptide (h, human; p, porcine)
N-Y4	N-terminal domain of Y4 receptor
Ni-NTA	nickel-nitrilotriacetic acid
OGP	octyl gluco pyranoside
OD	optical density
PAM	peptidylglycine α-amidating monooxygenase
PCR	polymerase chain reaction
PMX	Polymyxin
PP	pancreatic polypeptide (a, avian; b, bovine; h, human)
PYY	peptide YY
Re-LPS	LPS from deep rough mutant D31m4 of *E. coli*
RP-HPLC	reverse phase-high performance liquid chromatography
rpm	revolutions per minute
RT	room temperature
RU	response units; arbitrary unit in SPR
sCD14	Soluble CD14
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SPPS	solid phase peptide synthesis
SPR	surface plasmon resonance
TEMED	N,N,N',N'-**te**tra**m**ethyl**e**thylene**d**iamine
TEV	tobacco etch virus
TFA	trifluoroacetic acid
TOCSY	total correlation spectroscopy
Tris	tris-(hydroxymethyl)-aminomethane

TROSY	transverse relaxation optimized spectroscopy
TM	transmembrane
XEASY	**E**TH **a**utomated **s**pectroscop**y** on **X**-window
YUH	yeast ubiquitin hydrolase

Summary

This thesis dissertation involves two projects. The first project focuses on the interactions of bacterial lipopolysaccharide(LPS) with different types of polymyxins studied by NMR spectroscopy. The second project involves NMR structural studies of the N-terminal domains of the Y-receptors and the recombinant production of larger transmembrane domains from the neuropeptide Y4 receptor, as well as its structural and functional exploration using multi-dimensional NMR techniques. Lipopolysaccharide(LPS) forms the main constituent of the outer membrane of gram-negative bacteria. As a potent stimulator of the immune system, LPS is an important pathogenic factor for infection and inflammation playing a key role during the septic shock. The septic shock syndrome caused by the endotoxin still has an unacceptably high mortality rate and, owing to increasing numbers of antibiotic resistant strains, remains an ongoing threat throughout the world.

An understanding of the structural aspects of binding of LPS to its ligands in a membrane-mimicking environment would be very useful for the rational development of compounds with anti-endotoxin activity. We have focused for our study on two systems**:** 1. Binding of sCD14(1-152) to LPS and 2.Binding of different types of polymyxins to LPS. CD14 is a myelomonocytic differentiation antigen in mammals. It occurs in two forms a 50–55-kDa glycosylphosphatidyl-inositol (GPI)-anchored membrane protein (mCD14) and a monocyte or liver-derived soluble serum protein (sCD14) that lacks the membrane anchor. Both mCD14 and sCD14 are critical for LPS-dependent signal transduction, and sCD14 confers LPS sensitivity to cells which lacks mCD14. Mutant sCD14 containing only the N-terminal 152 amino acid residues were found to have and displays activity equivalent to the full length sCD14. Polymyxins are a group of cyclic polycationic polypeptides with a long hydrophobic tail isolated from various strains of *Bacillus polymyxa* and related species. They are selectively toxic for Gram-negative bacteria due to their specificity for lipopolysaccharide.

In the first stage, we started with the expression of sCD14(1-152) in *E.coli* and in the yeast *Pichia pastoris* expression system. Expression and purification of sCD14 was reported before, but the protein was highly aggregated precluding detailed structural studies by NMR. To increase solubility and retard aggregation we tested several fusion strategies and refolding trials to obtain properly folded sCD14(1-152). However, although some of the tested constructs were expressed at reasonable yields NMR experiments quickly revealed that they were either unfolded or highly aggregated. Since a crystal structure of the full-length protein was published during the studies we decided to not longer pursue that project.

Next, we investigated the structural properties of the LPS-Polymyxin (PMX) complex. NMR studies of polymyxins bound to LPS were published but no experimental data on the complex nor on the LPS-interacting residues were known. We present a study of interactions of PMX-B, -E and -M with LPS from the

deep rough mutant strain of *E.coli* in membrane-mimetic dodecylphosphocholine (DPC) micelles. This mutant of LPS is a so-called heptose-less mutants, in which the core, lipid-A, is only decorated by 2 Kdo (2-Keto-3-deoxyoctonate) moieties. First a method for efficient purification of biosynthetically produced LPS using RP-HPLC in a ternary solvent mixture was developed. Polymyxin-M(PMX-M) was biosynthetically produced from the *B.polymyxa* strain. The bacteria were grown on isotopically labeled medium, from which polymyxin-M was secreted into the medium and purified by chromatography. LPS was incorporated into a phospholipid micelles and its interaction with polymyxins was studied by heteronuclear NMR spectroscopy. The data enabled us to localize the interaction sites between LPS and polymyxins. Using simulated annealing protocol and restraints derived from intermolecular NOEs a model of the LPS–PMX-B complex was established, which was also in accordance with the chemical shift mapping data. The model was further refined by MD calculations, which included DPC micelles and explicit water. In the modelled complex the macrocycle of PMX-B is centered around the phosphate group at GlcN-B (N-acetylglucosamine-B) and additional contacts from polar side chains are formed to GlcN-A and Kdo-C, while hydrophobic side chains target the acyl region. In this study, we have synthesized and characterized all the constituents of the complex and the nature of the interacting moieties.

My second project involves structural studies of fragments of a human G-protein coupled receptor, in particular the Y4-receptor. Y-receptors are members of the rhodopsin-like G-protein coupled receptor (GPCR) family 1b. The Y-receptors are activated by neuropeptide Y, pancreatic polypeptide, and peptide YY, which belong to the family of neuroendocrine hormones. These receptors play an important role in the regulation of blood pressure, memory retention, food uptake and seizure. In this work, we focussed initially on structural studies of the extracellular N-terminal domains of the Y-receptors, and report on structural properties of the N-terminal domains of the human Y1-, Y2-, Y4-, and Y5- receptors. Synthetic routes for recombinant production of the polypeptides in isotopically-labeled form are described and compared to each other. The N-terminal domains from these Y receptors are fully unstructured in aqueous solution, as shown by measurements of the internal backbone dynamics. In contrast, in the presence of phospholipid micelles the N-terminal domain of the Y4 receptor (N-Y4) adopts a short α-helix in a segment mainly comprised of hydrophobic residues. N-Y1 is largely helical although remaining flexibility precludes a detailed structural analysis. N-Y5 is segregated into more structured and rather flexible regions, similarly to N-Y4. However, measurements of internal backbone dynamics revealed secondary structure to be less stable than in N-Y4. N-Y2 does not interact with the micelles and remains unstructured also in that environment.

In the second stage, on basis of structure and functional studies of N-Y4, we have extended our studies to longer transmembrane fragments, which contain two or more helices from the membrane domain, thus better mimicking the native properties of the receptor. In this study we describe about the expression and purification strategy for the production of double transmembrane fragments containing helices 4 and 5 (TM4-TM5 [I2(147)TM4-E2-TM5-I3(259)]) and helices 6 and 7 (TM6-TM7-C [I3(260)TM6-E3-TM7-C]) of the

Y4 receptor in isotopically enriched form in *E. coli.* Both the double transmembrane constructs were successfully expressed as Trp leader (TrpΔLE) fusion proteins, from which they were liberated by either cyanogen bromide (CNBr) cleavage (TM4-TM5) or by enzymatic cleavage(TM6-TM7-C) in the presence of detergents using C3 protease and purified by a combination of affinity chromatography and RP-HPLC. The cleavage strategy is an important point of concern because for many membrane proteins to be soluble at useful concentrations they require more acidic or more basic pH levels, high or low salt levels, or the presence of detergents or chaotropes. We tested using the N-terminus of Y4 receptor as a small fusion tag to favour the expression of double-transmembrane constructs. Accordingly, a N(Y4)-TM4-TM5 construct [(158)TM4-E2-TM5-I3(259)] was designed with a CNBr cleavage site to remove the N-terminal domain fusion if needed. This fusion strategy takes advantage of the small size of the N-terminal domain (41 aminoacids), which may allow direct NMR analysis without the need of removing the fusion protein. Two constructs were made, one, in which the I2 loop region before the TM4 region was included and one, in which this loop sequence was absent. [^{15}N,^{1}H]-HSQC spectra in detergent micelles revealed that the TM4-TM5 fragments in absence of the intracellular-2 (I2) loops form larger aggregates thus revealing the additional importance of loop residues for structural studies of the transmembrane segments.

Expression yields were optimized using different strains and growth temperatures as well as various induction parameters. The yield was found to be approximately 3-4 mg/L for TM4-TM5 fusion constructs and for TM6-TM7-C fusion protein 4-5 mg/L was obtained. The CNBr cleavage conditions were optimized, and cleavage efficency was found to be approximately 75-80%. The C3 protease cleavage, carried out in CHAPS or LDAO, still requires optimization of other buffer ingredients to maximize cleavage efficiency. In case of the TM4-TM5 construct, an extensive detergent screening using 15N-labelled protein that resulted in spectra of reasonable quality. Structures of these constructs in addition to the N4-TM1-TM2 fragment, previously studied in our group, will be an important step towards an structural understanding of the enitre receptor in detergent micelles.

Zusammenfassung

Diese Dissertation beinhaltet zwei Projekte. Das erste untersucht mittels NMR Spektroskopie die Wechselwirkung von bakteriellem Lipopolysaccharid (LPS) mit verschiedenen Arten von Polymyxinen. Das zweite widmet sich - ebenfalls durch Anwendung von mehrdimensionalen NMR-spektroskopischen Methoden - strukturellen und funktionellen Aspekten rekombinant hergestellter N-terminaler Domänen der Y-Rezeptoren und grösserer Transmembrandomänen des Y4-Rezeptors. LPS ist der Hauptbestandteil der äusseren Membran gram-negativer Bakterien. Als potenter Stimulator des Immunsystems stellt LPS einen wichtigen Faktor in der Infektion und Entzündung dar und spielt zudem eine wichtige Rolle beim septischen Schock. Die während des septischen Schocks durch das Endotoxin LPS hervorgerufenen Symptome haben immer noch eine inakzeptabel hohe Mortalitätsrate weltweit, insbesondere auch aufgrund einer wachsenden Anzahl antibiotikaresistenter Stämme.

Ein Verständnis der strukturellen Aspekte der Bindung von LPS an seine Liganden in einer membranmimetischen Umgebung wäre sehr nützlich für die rationale Entwicklung von Verbindungen mit anti-endotoxischer Aktivität. Für unsere Studien haben wir uns auf zwei Systeme konzentriert: 1. Bindung von sCD14(1-152) an LPS und 2. Bindung verschiedener Arten von Polymyxinen an LPS. CD14 ist ein myelomonozytisches Differentiationsantigen in Säugetieren. Es liegt sowohl in Form eines 50-55 kDa, via Glycosylphosphatidylinositol (GPI) verankerten Membranproteins (mCD14), als auch eines monozytischen, aus der Leber stammenden, löslichen Serumproteins (sCD14), welches keine Membranverankerung aufweist, vor. Sowohl mCD14 als auch sCD14 sind notwendig für LPS-abhängige Signalleitung. sCD14 ist dabei in der Lage Zellen, welchen mCD14 fehlt, sensitiv gegenüber LPS zu machen. Ein sCD14 Abkömmling, welcher nur aus den N-terminalen 152 Aminosäuren besteht, zeigt diesbezüglich die gleiche Aktivität wie vollständiges sCD14. Polymyxine sind eine Gruppe zyklischer, polykationischer Polypeptide mit einem grossen hydrophoben Teil und wurden aus verschiedenen Stämmen der Familie *Bacillus polymyxa* und artverwandter Spezies isoliert. Sie sind selektiv toxisch für gram-negative Bakterien aufgrund ihrer Spezifität gegenüber LPS.

In einer ersten Phase begannen wir mit der Expression von sCD14(1-152) in *E. coli* und in der Hefe *Pichia pastoris*. Über die Herstellung und Aufreinigung von sCD14 wurde bereits berichtet, aber das Protein lag in aggregierter Form vor, was strukturelle Untersuchungen mittels NMR verunmöglichte. Um die Löslichkeit zu erhöhen und die Aggregation zu verhindern und so korrekt gefaltetes sCD14(1-152) zu erhalten, haben wir mehrere Fusions- und Rückfaltungsstrategien ausprobiert. Aber obwohl einige der untersuchten Konstrukte mit guten Ausbeuten exprimiert werden konnten, zeigten NMR Experimente schnell, dass die so hergestellten Proteine entweder entfaltet oder aggregiert waren. Da zudem während

dieser Studien eine Kristallstruktur von sCD14 veröffentlicht wurde, haben wir uns entschlossen dieses Projekt nicht weiter zu verfolgen.

Als Nächstes haben wir die strukturellen Eigenschaften des LPS-Polymyxin Komplexes untersucht. NMR Studien zu diesem Thema wurden bereits andernorts durchgeführt, jedoch konnten keine Daten, welche die Komplexbildung erklärten, erhalten werden. Wir präsentieren eine Studie der Wechselwirkungen von Polymyxin-B, -E und -M (PMX-B, -E und -M) mit LPS der "deep rough" Mutante von *E. coli* in einer membranmimetischen Umgebung aus Dodecylphosphocholin (DPC) Mizellen. Diese Mutante produziert ein LPS, das heptoselos ist, und in welchem der Kern, genannt Lipid A, nur mit 2 Resten des Zuckers 2-Keto-3-desoxy-octonat (Kdo) versehen ist. Zuerst wurde eine Methode für die effiziente Reinigung von biosynthetisch hergestelltem LPS mittels RP-HPLC in einem ternären Lösungsmittelgemisch entwickelt. Polymyxin-M wurde im *B. polymyxa* Stamm - ebenfalls biosynthetisch - hergestellt. Die Bakterien wurden in isotopenmarkiertem Medium vermehrt. Polymyxin-M wurde dabei ins Medium sekretiert und konnte daraus chromatographisch gereinigt werden. LPS wurde in Phospholipidmizellen integriert und seine Interaktion mit den Polymyxinen mittels heteronuklearer NMR Spektroskopie untersucht. Die gewonnen Daten ermöglichten uns die Interaktionsstellen zwischen LPS und den Polymyxinen zu bestimmen. Durch Verwendung eines "simulated annealing" Protokolls und Distanzeinschränkungen aus intermolekularen NOEs konnte ein Modell des LPS-PMX-B Komplexes erstellt werden, welches in Einklang war mit anderen NMR Daten. Das Modell wurde durch molekulardynamische Simulationen, welche auch DPC Mizellen und Wassermolküle enthielten, weiter verfeinert. Im modellierten Komplex ist der PMX-B Makrozyklus um die Phosphatgruppen an N-Acetylglucosamin-B (GlcN-B) zentriert. Zusätzlich gehen polare Seitenketten weitere Interaktionen mit GlcN-A und Kdo-C ein, während sich hydrophobe Seitenketten an die Acylreste von LPS anlagern.

Das zweite Projekt war strukturellen Untersuchungen an Fragmenten von menschlichen G-Protein gekoppelten Rezeptoren (GPCRs) gewidmet, insbesondere dem Y4-Rezeptor. Y-Rezeptoren sind Mitglieder der rhodopsinähnlichen GPCR Familie 1b. Sie werden durch Neuropeptid Y, pankreatisches Polypeptid und Peptid YY aktiviert, welche alle zur Familie der neuroendokrinen Hormone gehören. Diese spielen wichtige Rollen in der Regulation des Blutdrucks, des Gedächtnisses, der Nahrungsaufnahme und im Entstehen von Schlaganfällen. In der vorliegenden Arbeit haben wir uns erst auf die N-terminalen Domänen der menschlichen Y1-, Y2-, Y4-, und Y5-Rezeptoren (N-Y1, N-Y2, N-Y4 und N-Y5) konzentriert. Synthetische Strategien für die rekombinante Produktion in isotopenmarkierter Form werden beschrieben und miteinander verglichen. Diese N-terminalen Domänen sind in wässriger Lösung komplett unstrukturiert wie Dynamikmessungen des Peptidrückgrats ergeben haben. In Anwesenheit von Phospholipidmizellen jedoch bildet sich in N-Y4 eine kurze α-Helix in einem Segment, welches hauptsächlich aus hydrophoben Aminosäuren besteht. N-Y1 ist unter diesen Bedingungen grösstenteils helikal. Eine hohe Flexibilität verunmöglicht jedoch eine detaillierte strukturelle Charakterisierung. N-Y5

kann in eine strukturierte und eine ziemlich flexible Region unterteilt werden, ähnlich N-Y4. Dynamikmessungen des Peptidrückgrats haben aber ergeben, dass die Sekundärstruktur in N-Y5 weniger stabil ist als in N-Y4. N-Y2 zeigt keine Interaktion mit Mizellen und bleibt auch in dieser Umgebung unstrukturiert.

Basierend auf den strukturellen und funktionellen Studien von N-Y4 haben wir unsere Bemühungen auf längere, transmembranäre Fragmente, welche zwei oder mehr Helices enthalten, ausgedehnt. Diese Fragmente kommen dem natürlichen Rezeptor aufgrund ihrer grösseren Länge näher. Wir beschreiben hier die Expression und Aufreinigung von Fragmenten aus Helices 4 und 5 (TM4-TM5 oder [I2(147)TM4-E2-TM5-I3(259)]) und Helices 6 und 7 (TM6-TM7-C oder [I3(260)TM6-E3-TM7-C]) des Y4-Rezeptors in isotopenmarkierter Form aus *E. coli* Bakterien. Beide Konstrukte wurden erfolgreich als Fusionsproteine mit der "Trp-leader" Sequenz (TrpΔLE) exprimiert, von welcher sie entweder mittels Cyanogenbromid (CNBr) Spaltung oder durch enzymatische Spaltung mittels der C3 Protease in Anwesenheit von Detergentien freigesetzt wurden. Gereinigt wurden die Fragmente mit einer Kombination aus Affinitätschromatographie und RP-HPLC. Die Wahl einer geeigneten Spaltungsstrategie ist ein kritischer Punkt, weil viele Membranproteinfragmente nur unter sehr harschen Bedingungen löslich sind. Wir haben unter anderem auch N-Y4 als einen kleinen Fusionspartner ausprobiert, welcher die Expression der Doppel-TM Konstrukte erhöhen sollte. So haben wir ein N(Y4)-TM4-TM5 Konstrukt ([(158)TM4-E2-TM5-I3(259)]) mit einer CNBr Spaltungsstelle, welche - falls gewünscht - die nachträgliche Abspaltung von N-Y4 erlaubt, erstellt. Diese Strategie nützt die geringe Grösse von N-Y4 (41 Aminosäuren) aus, welche auch eine direkte NMR Untersuchung ohne die vorherige Abspaltung des Fusionspartners ermöglicht. Zwei solcher Konstrukte wurden hergestellt: Eines, in welchem der I2-loop vor der TM4 Region enthalten war, und ein Anderes, in welchem diese Sequenz nicht enthalten war. [^{15}N,^{1}H]-HSQC Spektren in Mizellen zeigten, dass die TM4-TM5 Fragmente in Abwesenheit des I2-loops aggregierten. Die Anwesenheit dieses Loops war darum unerlässlich für NMR Studien.

Expressionsausbeuten wurden mittels verschiedener Stämme, Wachstumstemperaturen und Induktionsstrategien auf 3-4 mg/L (TM4-TM5) und 4-5 mg/L (TM6-TM7-C) optimiert. Die Effizienz der CNBr Spaltung konnte auf 75-80% optimiert werden. Die Spaltung mit der C3 Protease wurde in CHAPS oder LDAO Detergens durchgeführt und lässt noch Raum für Optimierung der Bedingungen. Für das TM4-TM5 Konstrukt wurde eine grosse Menge an Detergentien auf ihre Fähigkeit gute NMR Spektren herzugeben getestet. Strukturen von diesen Konstrukten werden einen wichtigen Schritt in die Richtung des Verständnisses des ganzen Rezeptors in Detergensmizellen darstellen.

Chapter 1

1.0 Introduction

Proteins and enzymes that bind carbohydrates are present in large numbers in all living cells and are involved in a myriad of important biological functions. This abundance is primarily due to the fact that carbohydrates derived from carbon dioxide fixation constitute the bulk of the organic matter on earth, and together with their various derivatives and polymeric forms, are utilized for many essential purposes in the cell. Sugars are the central energy source for mechanical work and chemical reactions in living cells. Phosphate derivatives of monosaccharides are important in these energy transformations, and several sugar derivatives, particularly adenosine triphosphate, serve key roles in the storage and transfer of energy. The uptake of sugars into the cell and their subsequent utilization require transport proteins and a whole assemblage of enzymes. In addition, carbohydrates in the form of polysaccharides are used for fuel storage and as structural elements[1].

Protein–carbohydrate interactions are the basis of numerous biological processes, both normal and pathological ones. They play crucial roles in many biological mechanisms, including enzymatic synthesis and degradation of oligo- and polysaccharides, intracellular sorting of glycoconjugates, transport of carbohydrates into living cells and of their derivatives into subcellular organelles, the immunological response to carbohydrate antigens, and migration of leukocytes to sites of inflammation (so-called homing). These interactions also play a key role in a variety of cell adhesion phenomena, among them are the attachment of parasites, fungi, bacteria, and viruses to host cells, the first step in the initiation of infection. In these processes, they essentially depend on the molecular recognition of specific carbohydrates by proteins[2]. In this context, the determination of the structural and conformational factors, which govern the molecular recognition of these biomolecules, as well as the knowledge of the physicochemical features of these processes, is of paramount importance. Specificity of these interactions is mediated by various tissue dependent receptors. Although extensive biochemical work has been done in that field to determine which units of the oligosaccharides are important for the recognition process, little structural data on the orientation of these with respect to the membrane surface and about the steric

requirements for their interaction with ligands is available. Importantly, in the vicinity of the water–membrane interfacial area certain properties are dramatically different when compared to them in bulk aqueous solution. It will be therefore interesting to investigate, whether membrane–coupled carbohydrates adopt specific conformations that improve binding their targets in comparison to the non-coupled species.

A deeper understanding of these interactions at the molecular level will enable the development of novel, effective and highly selective therapeutics. Binding and recognition phenomena are essential processes by which the body exerts control over complex biological functions. Important targets for therapeutic intervention are the binding processes mediated through multivalent protein-carbohydrate interactions, such as the interactions of bacterial toxins with cell-surface receptors. Evidently, the structure and dynamics of the interacting entities are crucial for recognition. In this context, NMR has become a major tool for scrutinizing the conformational behavior and interaction properties of oligosaccharides. The development of new NMR techniques and access to stronger magnetic fields and to new technologies (e.g. cryoprobes and better electronics) are expanding the field in ways that were difficult to foresee a few years ago.

1.1 Chemistry of protein–carbohydrate interactions

A well-known characteristic of protein–carbohydrate interactions is the low affinity of binding, with dissociation constants usually in the millimolar range. Stronger binding or enhanced inhibition is often achieved by the use of multiple interactions by multivalent carbohydrates by a phenomenon called cluster effect[3]. These interactions are driven by a favorable enthalpy (heat is released). However, the favorable enthalpy is counteracted by an unfavorable entropy (degree of order) term that might stem from restricted carbohydrate flexibility. Enthalpy–entropy compensations are a characteristic of weak chemical interactions of many biological processes[4]. They comprise primarily a network of hydrogen bonds and hydrophobic interactions; in some cases electrostatic interactions (or ion pairing) and coordination with metal ions also play a role[5, 6].

1.1.1 Hydrogen bonds

Hydrogen bonds that are directional confer specificity to protein–carbohydrate interactions, as well as contributing to their affinity[7]. The hydrogen bonds are formed between carbohydrate hydroxyl groups and NH groups, hydroxyls, and oxygen atoms of the protein most frequently of aspartic acid, aparagine, glutamic acid, glutamine, arginine and serine residues. When each of two adjacent hydroxyls of a monosaccharide interacts with a different atom of the same amino acid (e.g., the two oxygens of the carboxylate of glutamic or aspartic acid), they form bidentate hydrogen bonds[7]. Such bonds are quite common in protein-carbohydrate complexes. A different kind of hydrogen-bond characteristic for such complexes is the cooperative bond, in which the hydroxyl group acts simultaneously as donor and acceptor. Polar–polar interactions are actually stronger than in the corresponding protein–protein interactions. Three major factors responsible for this are (1) higher surface density of hydrogen bonds; (2) better hydrogen-bonding geometry; (3) larger proportion of hydrogen bonds involving charged groups.

1.1.2 Hydrophobic interactions

Even though carbohydrates are highly polar molecules, the steric disposition of hydroxyl groups creates hydrophobic patches on sugar surfaces that can form contacts with hydrophobic regions in the protein molecules[8]. One widely occurring interaction of this kind is the stacking of a monosaccharide on a side chain of an aromatic amino acid, which is due to the presence of partial positive charges on the aliphatic protons on one face of a hexopyranose ring and a partial negative charge from the π electrons of the aromatic system[9].

1.1.3 Electrostatic interactions

Most saccharides are uncharged, and therefore ionic, *i.e.*, charge–charge interactions are less common however these interactions occur between negatively charged lipopolysaccharide (LPS) and cationic antimicrobial peptides like polymyxin[10]. The importance of electrostatic interactions for polymyxin-LPS binding also is evident from its inhibition by divalent cations and high ionic strength[11].

1.1.4 Role of water

As the human body has a water content of 70%, it is not surprising that water plays an essential role in both protein–protein and protein–carbohydrate interactions[12]. Due to its small size and ability to serve as both a hydrogen bond donor and acceptor it can act as a bridge in protein ligand interactions. Generally, complementary surfaces with well-matched polar interactions cause the escape of receptor/ligand-bound water molecules to the bulk, a reaction that favors the binding. In contrast, poorly matched surfaces result in incomplete desolvation of the molecules involved, and a low affinity binding[13]. Thus the individual changes in enthalpy and entropy might be substantial, yet the change in free energy of binding ($\Delta G°$) between a receptor and different ligands is relatively small. Consequently, increasing the affinity (i.e. increasing $\Delta G°$) of specific protein–carbohydrate interactions remains a complicated task.

1.2 Tools for analyzing carbohydrate protein interactions

To understand the protein ligand interactions, we should know in detail the relationship between function and molecular recognition. Structure, kinetics, energetics, dynamics, stoichiometry and dissociation or association equilibrium constants have to be defined as accurately as possible. Here I will discuss in detail the NMR methodologies used and very briefly the other major techniques like X-Ray crystallography, molecular dynamic simulations, fluorescence spectroscopy, surface plasmon resonance, isothermal titration calorimetry and differential scanning calorimetry. There are many factors, which have to be taken into account, when applying a technique for the analysis of protein-ligand interactions.

1.2.1 NMR approaches to study carbohydrate–protein interactions

High resolution NMR spectroscopy has proved to be a very useful technique for studying interactions between proteins and other molecules, which are of crucial importance in many molecular recognition processes in biology. It directly yields valuable experimental data on biologically important protein–ligand complexes.

1. Qualitative and quantitative binding affinity: Changes in NMR parameters allow us to detect and quantitatively determine binding affinities of potential ligands.

2. Determination of binding site(s): Based on changes in the assigned NMR signals of a protein upon ligand addition, a ligand binding site on the protein can be mapped.
3. Conformational information: The conformation of the protein and/or ligand in the complex can be determined and compared to their free states.
4. Dynamic information: In addition to the static structure, local dynamics can be measured for the complex and the free components to yield a more accurate picture of the intermolecular interactions.

NMR investigations allow us to deduce (*a*) structural information on the protein residues involved in the interaction; (*b*) information on the bound carbohydrate; and (*c*) information on the entire protein–carbohydrate complex. The first method observes the changes in chemical shift of the target resonances upon ligand binding through shift mapping. It requires milligram quantities of ^{15}N- or ^{13}C-labeled target, and resolved, heteronuclear, two-dimensional NMR correlations (e.g., HSQC, heteronuclear single-quantum coherence) for monitoring chemical shift changes. If the target is large (>30 kDa), extensive resonance overlap might prevent the complete and unambiguous interpretation of the spectra. A distinct disadvantage of the method is the necessity of a complete resonance assignment of the target or at least of the binding site of interest. The second methodology, which relies on the observation of the ligand resonances, is exclusively applicable to ligands in the low-affinity range (high micromolar to millimolar range). It is commonly used when either the target is too large to be observed by NMR (>50 kDa), or it is not available with the desired ^{15}N- or ^{13}C-labeling pattern from *in vitro* production, or it aggregates in solution at higher concentrations. All these scenarios prevent the collection of highly resolved spectra of the target.

A number of NMR variables can be monitored in this screening assay: (*a*) line broadening, (*b*) change of the NOE (nuclear Overhauser effect) from positive to negative (transferred NOE), (*c*) intermolecular magnetization transfer, (*d*) restricted ligand diffusion, and (*e*) change in the relaxation properties of the ligand due to interaction with spin labels[14]. From a kinetic point of view, it could also be possible to say something about the ligand exchange timescale and the size of the complex (relaxation measurements)[15]. Thus a wide variety of NMR methods can be exploited

to characterize macromolecular complexes (in particular, protein-protein, -DNA, -RNA, -sugar, -lipid, -drug, etc).

The technique of choice will depend upon several factors[16], which are as follows

• size and type of molecules.
• equilibrium binding constant (strong vs. weak).
• kinetics of binding (fast vs. slow exchange).
• stoichiometry and symmetry.
• available nuclei (^{1}H, ^{13}C, ^{15}N, ^{19}F, ^{31}P, ^{113}Cd, ...)

1.2.2 NMR parameters that yield information on protein ligand interactions

1.2.2.1 Chemical shift

Chemical shifts are exquisitely sensitive markers of the ligand charge state environment at the bound site. A qualitative interpretation of changes in chemical shifts on binding provide significant insight into the nature of the binding site even though it is not possible to construct an accurate model of the binding site using this knowledge. ^{1}H and ^{13}C have been the nuclei most widely studied, and the dependence of these shifts in proteins on secondary structure is well established[17, 18]. Nitrogen chemical shifts of backbone amides are sensitive to many factors like hydrogen bonding of corresponding protons, which makes them difficult to interpret. They are used only to increase the dimensionality of the spectra to allow assigning the individual amino acids.

1.2.2.2 Relaxation parameters

Transverse and longitudinal relaxation of proton spins is mainly prompted by dipolar interactions between proximate spins. The relaxation efficiency depends on the distance separation of the interactions and on the reorientation time of the internuclear vector. Thus the spin-lattice relaxation time T1 and the spin-spin relaxation time T2 are dependent on overall and internal molecular mobility and hence provide sensitive probes of binding. Binding of the ligand to the target leads to dramatic changes in overall tumbling for the bound fraction of ligand. T2 values traditionally have been

more often utilized in the study of macromolecule-ligand interactions than T1 values. Binding of ligand leads to significant changes in linewidth of the ligand signals.

1.2.2.3 NOEs

The nuclear Overhauser effect (NOE) is defined as the change in intensity of a certain resonance after perturbation of the population of a spin in close spatial proximity (by saturation or inversion). The interaction through space of the dipoles of neighboring nuclei, or cross-relaxation, gives rise to the NOE. The cross relaxation rate σ, which depends on the rate of molecular tumbling, governs the magnitude and sign of the NOE[19]. Low- or medium-molecular-weight molecules (MW<1-2 kDa) have a short correlation time τ_c and, as a consequence, such molecules exhibit positive NOEs, no NOEs, or very small negative NOEs depending on their molecular weight, shape, and the field strength. Large molecules, however, exhibit strongly negative NOEs.

NOEs are the most useful NMR parameter for the analysis of protein-ligand complexes. By establishing the proximity in space of two nuclei (via NOEs), it is possible to deduce conformational information (in the case of intramolecular NOEs) or locate a binding site (via intermolecular NOEs). Although NOEs have been used widely for the former application in the past, for conformational and structural studies of both small molecules[20] and proteins[21], it has been only in the last few years that major advances have been made in the latter application. The detection of significant number of intermolecular NOEs for tightly bound protein-ligand complexes has been driven by the development of isotope editing techniques that have rendered the assignment of such NOEs possible.

When a small molecule (ligand) is bound to a large-molecular weight protein (the protein receptor molecule) it behaves as a part of the large molecule and adopts the corresponding NOE behavior, that is, it shows strong negative NOEs, so-called trNOEs. These trNOEs reflect the conformation of the bound ligand. Binding of a ligand to a receptor protein can thus easily be distinguished by looking at the sign and size of the observed NOEs.

1.2.2.4 Magnetization transfer rates

Information about exchange processes can be provided by monitoring the changes in peak intensity in a NMR spectrum. In particular, the use of saturation transfer or inversion transfer techniques allows information about the kinetics of exchanging systems to be deduced. The rate of loss of NH signal intensity following dissolution of the sample in D_2O provides a measure of the degree of protection of the proton from solvent revealing its involvement in hydrogen bonding underlying secondary structure stabilization.

1.2.2.5 Aspects of binding affinity and timescale for detection of NMR parameters

The type of experiment most appropriate for the system is dependent on the exchange rates between the ligand and protein relative to the NMR time scale[22]. The thermodynamic and kinetic properties (dissociation constant and on/off rates) of the protein-ligand interactions influence the outcome of the NMR parameters such as the chemical shift, the relaxation rates of the ligand, and the Nuclear Overhauser effect (NOEs). On the NMR time scale, exchange regimes are broadly classified as slow, intermediate and fast with the regime being related to the equilibrium dissociation constant, K_D. The binding of a ligand L to an enzyme E can be considered a two site second-order exchange.

$$E + L \leftrightharpoons EL$$

The above equilibrium is described by the dissociation constant, $K_D = k_{off} / k_{on}$, where k_{off} is the rate of dissociation (off–rate) of the ligand from the protein and k_{on} is the rate constant for association (on–rate) of the ligand with the protein. The off- and on-rate define the lifetime τ of the particular state. The lifetime of a NMR nucleus in the ligand molecule in state EL is given by $\tau_{EL} = 1/k_{off}$, and the lifetime in the state L by $\tau_L = 1/ (k_{on} [E])$. For a nucleus in the protein the lifetime in the state EL is given by $\tau_{EL} = 1/k_{off}$ and the lifetime in the state E by $\tau_E = 1 /(k_{on} [L])$. E and L are respectively, the concentration of the free enzyme and the ligand and [EL] the concentration of the complex. A single lifetime is used to characterize the process defined by the equation. For ligand

$$1/\tau = 1 /\tau_{EL} + 1/\tau_L = k_{off} (1+ P_{EL} /P_L)$$

where P_{EL} = [EL]/[L_{tot}] and P_L = [L] / [L_{tot}], respectively, the fraction of the ligand in the bound and free states, and L_{tot} = [EL] + [L]. It is important to note that the τ and hence the actual appearance of the spectrum depends on the ligand concentration.

For a typical millimolar protein concentration, depending on the type of interaction we have 1) strong binding with dissociation constants K_D in the sub micromolar range 2) moderate binding with micromolar K_D values and 3) weak interactions with millimolar dissociation constants. For strongly binding ligands, both the protein and ligand exist almost exclusively (>95%) in the bound state and the complex behaves as a single stable molecule. The spectra will not show any signs of the free species or of exchange between free and bound species. In the moderately binding case, under stoichiometric conditions both protein and ligand will exist as bound and free species in significant percentages. Depending on the kinetic stability, exchange between the free and bound species might also affect the NMR spectra. However this can be driven to a completely complexed state by adding 10 to 100 fold excess of the ligand. A lot of information about the complex can be gained from NMR measurements (usually involving isotope filtering) e.g. indications of conformational changes upon complexation and molecular regions involved in binding. Assuming that the on–rate k_{on} is diffusion limited, one can estimate the range of complex lifetimes for different affinities (Table1).

:

K_D	**k_{off}**	**Life time of the complex**
1 nM	0.1 s^{-1}	10 s
1 uM	100 s^{-1}	10 ms
1 mM	10^5 s^{-1}	0.1 ms

Table 1: Lifetime of protein-ligand complexes with different affinities

Parameters	Typical range (Hz)	Typical difference between free and bound states (Hz)	Lifetime at intermediate exchange
Chemical shift	0 - 10000	100 - 1000	1 - 10 ms
Scalar couplings	0 - 200	1 - 10	0.1 – 1s
NOE (build–up rates)	0 - 5	0.1 - 5	0.2 – 10 s
Relaxation times T1 T2	0.1 – 10 1 - 100	0.1 – 10 1 - 50	0.1 – 10 s 0.02 – 1s

Table 2: Exchange time scale for NMR parameters

In order to measure NMR parameters (represented by a frequency or rate) separately for each state, its lifetime has to be significantly longer than the inverse of the differences of this parameter in the two states (slow exchange). If it is much shorter than the intermediate exchange regime then only a single set of signals will be observed, reflecting an average over both the free and the complexed species, weighted by their relative concentration (fast exchange).

Slow exchange $1/\tau << 2\pi|\nu_L - \nu_{EL}|$

Intermediate exchange $1/\tau \approx 2\pi|\nu_L - \nu_{EL}|$

Fast exchange $1/\tau >> 2\pi|\nu_L - \nu_{EL}|$

The fast and slow exchange regimes are not fixed, but depend on the NMR parameters under consideration and also on their specific differences in the two species. Except for high affinity ligands with nanomolar K_D values, lifetimes usually range in the millisecond range for most complexes. From the Table 2, it can be seen that the parameters least sensitive to averaging are the chemical shifts and to a lesser degree, T2 relaxation rates. NMR parameters which serve as measure of binding and can be obtained easily with high sensitivity are chemical-shift changes, changes in relaxation times, changes of diffusion constants, changes of NOEs, or exchange of saturation. A number of criteria have to be considered to choose the right method to identify and characterize protein-ligand interactions by NMR spectroscopy.

1.2.3 Experiments based on chemical shift perturbation

NMR spectroscopy provides detailed information on the chemical surroundings of a given nucleus through chemical shifts and therefore, NMR titration experiments may provide an adequate means to analyze sugar-induced perturbations of proteins and *vice versa*. For the protein, nuclei located in the protein-binding pocket usually show the largest effects. Because of the signal overlap in the 1D spectra of proteins changes in chemical shift are most conveniently monitored by heteronuclear correlation experiments. This method is referred to as chemical shift mapping[23]. The HSQC based methods are presently limited to macromolecular complexes with a molecular weight below 30 or 100 kDa, depending on whether complete deuteration of the protein (in combination with TROSY) is used or not. Another limitation is the availability of larger quantities of isotopically labeled protein. For ligand protons because of low molecular weight, spectra are less complex, with less overlapping of signals. Therefore, changes of ligand proton resonance characterize binding. One of the most common applications is the determination of dissociation constants from titration experiments, but more sensitive parameters (relaxation and diffusion rates) may be used for ligand-based NMR screening approaches.

1.2.4 Saturation transfer difference (STD) nuclear magnetic resonance experiments to detect ligand binding

The STD NMR method is capable of identifying the binding epitope of a ligand when bound to its receptor protein[24]. In this experiment, resonances of the protein are selectively saturated and intramolecular NOE effects build up rapidly and are transferred throughout the protein by spin diffusion and transfer of saturation from the protein to a bound ligand by intermolecular $^{1}H–^{1}H$ cross relaxation. Ligand protons that are in close contact with the receptor protein receive a higher degree of saturation, and as a result stronger STD NMR signals can be observed. Protons that are either less or not involved in the binding process reveal no STD NMR signals. In the STD NMR experiment signal attenuation of ligand signals in comparison of two separate experiments is observed, one recorded with saturation of target resonances (mostly the methyl hump of the protein is saturated) and the second one in which irradiation is performed far off-resonance (on-resonance experiment and off-resonance experiment). Subtraction of the *on*-resonance and *off*-resonance spectrum results in the final

saturation transfer difference (STD) NMR spectrum showing only signals from ligands that have binding affinity (Figure1). If interaction of non-labile protons is the focus, STD NMR experiments can be carried out in 100% D_2O, eliminating problems arising from the dominant water signal. The WATERGATE solvent suppression sequence or the excitation sculpting method can be used to suppress the H_2O signal in samples where labile protons need to be observed.

STD NMR spectroscopy can be utilised to detect binding ligands with K_D values in the 10^{-2} to 10^{-8} M range[25]. High-affinity ligands that undergo slow chemical exchange typically reside longer within the protein-binding site and are thus not detectable by STD NMR spectroscopy. Temperature can significantly affect STD NMR signals intensities and it is recommended to acquire a series of STD NMR spectra at different temperatures[26]. Saturation frequency (*on*-resonance) has to be set to a value where only protein resonances and no ligand signals are located. It is recommended that the protein saturation frequency is placed away at >700 Hz from the closest ligand resonance.

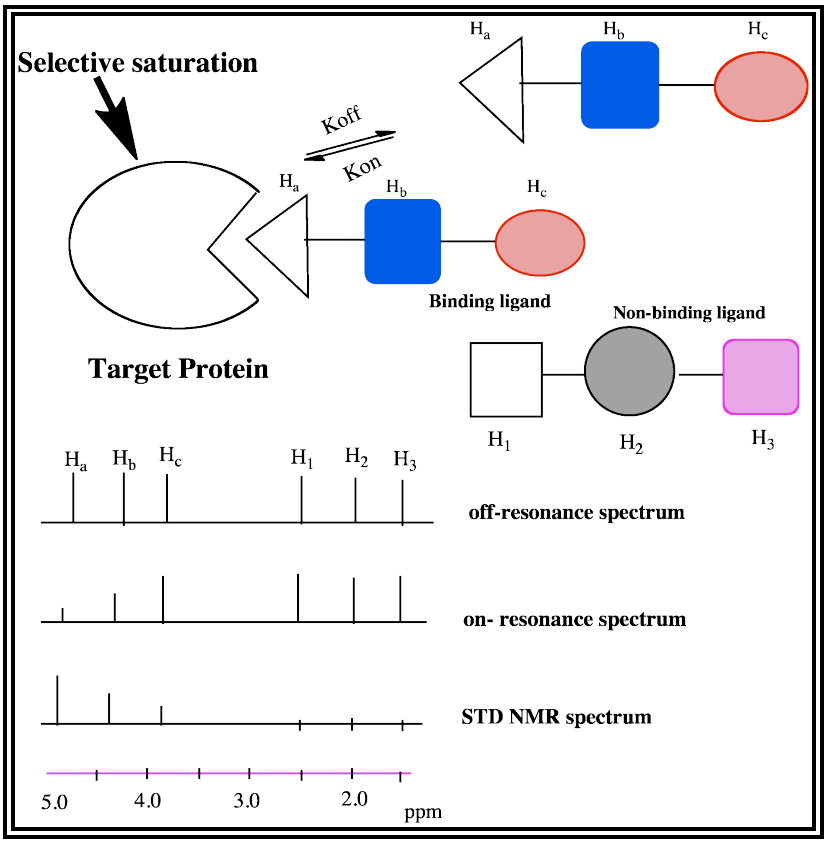


Figure 1: Scheme showing principle of the STD NMR method.

In contrast to transferred NOE experiments STD experiments do not critically depend on the ligand to protein ratio, as signal intensity that does not result from the saturation transfer process from the protein to the ligand is eliminated and the spectra always reflect purely the bound state. But usually a ratio of 1:100 (protein: ligand) is a good choice, as it prevents ligands that have already received saturation, from re-entering the protein-binding site. One major advantage of the new STD techniques is that the STD method may be combined with any NMR spectroscopy such as STD TOCSY or STD HSQC[24, 27] and only small amounts of native protein are required and no expensive protein labelling is necessary.

1.2.5 Water LOGSY: water-ligand observation with gradient spectroscopy

Water LOGSY is a variant of STD NMR spectroscopy which utilizes the bound water at the protein ligand interfaces[28]. It also relies on the transfer of magnetization between ligand and target. The transfer path is not direct (as in STD NMR) but uses H_2O molecules in the binding pocket as an intermediate magnetization pool. Thus, the sample solution must consist predominantly of H_2O with only a small quantity of D_2O for locking. Because the sign of the magnetization transfer signals via the nuclear Overhauser effect (NOE) is dependent on the molecular weight of the interacting molecules under investigation, a signal stemming from transfer between the water and the ligand (both of low molecular weight) will have a positive sign in the Water LOGSY spectrum. To eliminate artifacts like radiation damping pulsed field gradients were employed for proper water suppression. The high sensitivity of the technique reduces the amount of biomolecule and ligands needed for the screening.

1.2.6 Transferred intramolecular nuclear overhauser effects (trNOEs)

The observation of transferred NOE is possible for ligand–protein complexes that are characterized by dissociation constants K_D that are in the μM to mM range and receptor sizes in the range of 60–350 kDa. When ligand molecules bind to receptor proteins the NOEs undergo drastic changes leading to the observation of transferred NOEs (trNOEs). Too tight binding places the systems outside the range of fast exchange on the NMR time scale and no transferred NOEs are observed as the

exchange process between the free and bound forms of the ligand is not fast enough to ensure proper averaging of the NMR observables; on the other hand, the population of the bound ligand for weaker complexes becomes too low to observe a change in the NOE sign[29].

The amplitude and sign of NOEs depend on the rates of molecular motion (overall tumbling and internal mobility), which are usually characterized by an effective motional correlation time τ_c. The observation of trNOEs relies on different tumbling times τ_c of free and bound molecules. When a small molecule (ligand) is bound to a large-molecular weight protein (the protein receptor molecule) it behaves as a part of the large molecule and adopts the corresponding NOE behavior, that is, it shows strong negative NOEs, so-called trNOEs (Figure 2). These trNOEs reflect the bound conformation of the ligand. Binding of a ligand to a receptor protein can thus easily be distinguished by looking at the sign and size of the observed NOEs. Furthermore, the discrimination between trNOEs originating from the bound state and NOEs of the ligand in solution can also be achieved by the build-up rate, that is, the time required to achieve maximum intensity, which for trNOEs is in the range of 50 to 100 ms, whereas for nonbinding molecules it is four- to ten-times as long[25].

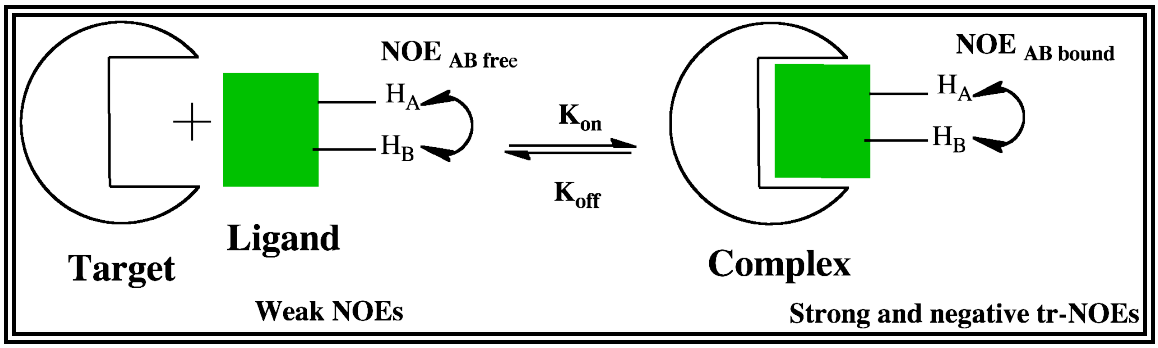


Figure 2: NMR measurements to detect carbohydrate binding to proteins by observation of transferred NOEs, in the ligand resonances.

The ligand: target concentration ratio [L]: [T] plays a fundamental role in the observation of transferred-NOE effects. The optimal [L]: [T] concentration ratio depends on the size of the target and on the kinetic constants involved. It is advisable to work at low protein concentrations and adjust the NOESY mixing time to obtain a good signal-to-noise ratio[14]. In many cases, this optimum is at a molar ratio of

carbohydrate to protein between 10 and 50 to 1. The effect of binding kinetics plays a role towards the sensitivity and thus the quality of the attained spectra. Intermolecular cross-relaxation between ligand and protein cannot be neglected and can lead to serious errors in the structure of the bound conformation of the ligand if derived from trNOE intensities, because of spin diffusion effects. Several techniques like QUIET-BAND-NOESY and QUIET-ET-NOESY have been developed to suppress spin diffusion[30], [31]. Intermolecular NOEs between protein and ligand are sometimes desirable because it allows detecting the orientation of the ligand in the protein binding pockets. Another complication of trNOE is the possibility of non specific interaction of the ligand[32]. Specificity of the interaction can be best demonstrated by a competition experiment with a strong binder.

1.2.6.1 The transferred ROESY experiment

It is an alternative method observing distances through space between nuclei in a molecule. In many cases the spin-diffusion effects can lead to the observation of 'false' trNOEs that do not reflect direct dipolar contacts between carbohydrate-ligand protons, but originate from magnetization transfer via protons attached to amino acids in the binding site of the protein receptor or via other ligand protons[33, 34]. It was found that transferred NOEs in the rotating frame, trROEs, are well suited to identify such spin diffusion pathways. In this experiment, spin-diffusion (three spin) effects appear as positive cross peaks and therefore, the application of this experiment permits one to distinguish direct from indirect enhancements, and thus complements those measured under regular conditions, providing conformational information which is less contaminated by artifacts.

1.2.7 Isotope filtering/editing using labeled ligands

Isotope filtering/editing NMR techniques make use of differential isotopic labeling to simplify spectra and thus more easily extract information from complex systems. The basic idea is to selectively observe the subspectra of the labeled or unlabelled components only (Figure 3). Isotope editing means selecting the protons, which are directly, bound to an NMR active heteronucleus (^{1}H - ^{13}C, ^{1}H – ^{15}N). Isotope filtering suppresses the signals leaving only the signals from the proton bound to the NMR inactive heteronuclei (^{1}H -^{12}C, ^{1}H-^{14}N**)** as well as protons not bound to carbon and nitrogen at all (mainly –SH , -OH). These methods rely on $^{1}J_{H, X}$ coupling, since these

are i) relatively large (requiring only short coupling evolution delays) and ii) fairly uniform (allowing exact tuning of delays). In these experiments the species actually observed is the complex itself. This is the important difference from trNOE or saturation techniques, where the existence of equilibrium between the bound and free species and a certain rate of exchange between them is essential. So the other general conditions for these experiments are identical to those required for standard protein NMR techniques. The systems under study should be in the strong binding regime to prevent the signals from significant concentration of the free species (in trNOEs intermediate affinity is preferred).

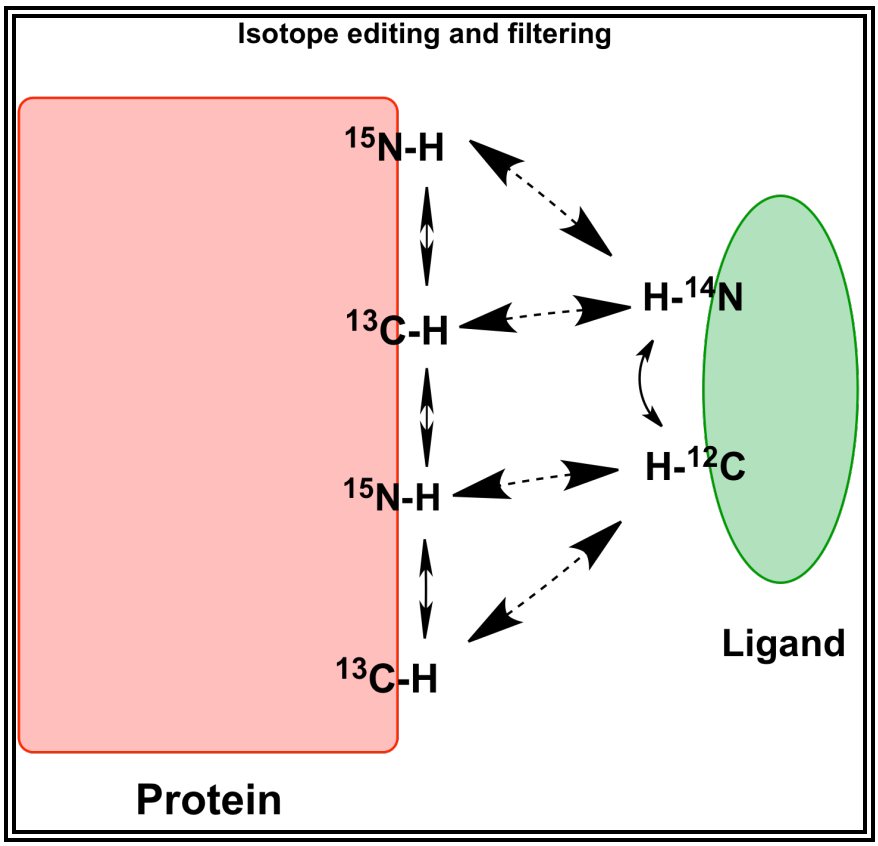


Figure 3: Schematic of isotope-filtered and -edited ^{1}H NMR as applied to a protein –ligand complex. In this example the protein is labelled uniformly with ^{15}N and ^{13}C, while the ligand is at natural isotopic abundance. Correlations within the labelled protein may be selectively observed using isotope-*edited* experiments (thin arrows). Correlations within the unlabelled ligand may be selectively observed using isotope-*filtered* experiments (curved arrow). Intermolecular NOE correlations between macromolecule and ligand may be selectively recorded using experiments that are isotope-filtered with respect to one proton dimension and isotope-edited or -separated with respect to the other (dashed arrows).

The main labeling strategy for these experiments are 1) One component is isotopically enriched to a reasonable extent and the other contains unlabelled species. 2) One component is practically 100% enriched with NMR active nuclei isotope, while the other component contains this isotope only to a much lower degree. In NMR studies of protein ligand interactions, isotope filters are an important concept for simplifying spectra of the systems otherwise too complex for an analysis. In this way, intra- and intermolecular NOEs involving ligand protons directly attached to ^{13}C and ^{15}N can be selectively detected. Only one component is labeled with stable NMR–active isotope (usually ^{15}N and /or ^{13}C) and the other one used in natural isotopic abundance. The NMR–active isotope can then be used for signal assignment via 2D or 3D heteronuclear correlation spectra and for improved resolution by adding a third dimension to 2D ^{1}H TOCSY or NOESY experiments. Since these pulse sequences require ^{15}N or ^{13}C spins, only the labeled part of the complex will show signals.

1.2.8 Spin labelling

A spin label (SL) is an organic molecule, which possesses an unpaired electron, usually on a nitrogen atom, and has the ability to bind to another molecule. The application of spin labels to NMR spectroscopy was based on the increase of relaxation rates of neighbouring protons caused by the paramagnetic center, the so-called paramagnetic relaxation enhancement (PRE)[35, 36]. The spectral perturbations caused by the electron spin labels decrease rapidly with increasing distance from the paramagnetic center (prop. r^{-6}). Both the increased proton relaxation rates and the altered ^{1}H chemical shifts can be evaluated to determine the distance of the perturbed nuclear spins from the unpaired electron[37]. It can be used to measure distances up to 20 Å which are too long to be measured by NOE experiments, or to study the dynamics of peptides or proteins by sampling all conformations that lead to short proton-radical distances.

The paramagnetic relaxation contribution to the line widths can be distinguished from chemical-exchange broadening by a control experiment in which the nitroxide spin label is reduced with ascorbic acid[36]. An additional consequence of paramagnetism is partial alignment of the paramagnetic molecule in the magnetic field, due to its anisotropic magnetic susceptibility[38]. Hence dipolar couplings do not

average out and residual dipolar couplings can be measured which provide long–range constraints for structure determination.

	Chemical shift perturbation	**STD NMR**	**Water-LOGSY**	**trNOEs**	**Isotope editing**	**Spin labelling**
Large protein (>30 kDa)	Limited	Yes	Yes	Yes	No	Yes
small protein (<10 kDa)	Yes	No	No	No	Yes	Yes
Isotope-labeled protein required	Yes	No	No	No	Yes	Yes
Binding epitope on ligand/protein	Protein if it is assigned	Ligand	Ligand	Ligand	Protein-ligand complex	None
Identification of Ligand	No	Yes	Yes but in the water contact surface	Yes	Yes	Yes
K_D tight binding	No limit	100 pM	100 pM	1 nM	~ 100 nM	100 pM
K_D weak binding	~ 1mM	~ 10 mM	~ 10 mM	~1 mM	~ 1 mM	~ 10 mM
Application	chemical-shift perturbation of resonances of the target	identifies compounds that bind weakly	identifies compounds that bind by using water-mediated NOEs	interaction of binders with the target	to study the protein – ligand complex	highly sensitive detection of binding Fragments.
Comment	robust method	robust method	good for very hydrophilic targets and/or ligands	good for screening ligands	Relatively insensitive	sensitive method

Table 3: A brief outline of NMR spectroscopy techniques for the identification and characterization of binding of ligands to proteins. Modified from reference 25.

1.3 Other methods to investigate protein-carbohydrate interactions

1.3.1 Molecular dynamics simulations

In many cases, not all of the required information can be obtained directly by experimental studies, so theoretical molecular modeling is usually required to

supplement experimental data, both in solution and in the solid state. Different modeling protocols have been proposed to locate protein binding sites[39]. Some of them can be used to calculate the interaction energies between ligand and receptor.

Other methods are designed to explore systematically all the positions and orientations that the sugar may adopt within the binding site[40]. Molecular docking increasingly becomes a valuable tool for exploring protein ligand interactions. It calculates the binding geometry and binding free energy of the complex if the spatial structure of protein and ligand is provided. Starting with the spatial structures of the binding molecules, a structure-generation step produces a great number of putative complex structures. These many complexes are then filtered by suitable fast filters or scoring functions in order to reduce the number of candidates. The top-scored structures are then evaluated energetically with an accurate energy function, thus yielding approximations of the binding mode and binding free energy of the complex. Thus docking schemes try and find local minima, while scoring functions try and look at the interactions from a global perspective.

The important features of protein carbohydrate interactions, which should be considered for molecular modeling approaches are 1) the large number of freely rotatable polar hydroxyl groups which leads to increased hydrogen bonding and strong electrostatic interactions, 2) the geometric complementarity of the ligand and the binding pocket in protein is less pronounced, because the sugar binding regions are often shallow grooves[41], 3) the glycosidic bonds of the carbohydrates are very flexible without pronounced rotameric states, resulting in a huge conformational space. In addition, due to the polar nature of the many hydroxyl groups their interactions are often dominated by solvation effects.

1.3.2 X-Ray crystallography

Crystal structures of protein-ligand complexes provide a detailed view of their spatial arrangement and interaction. X-ray crystallographic analysis has the potential advantage over NMR and theoretical methods that it can provide a complete oligosaccharide conformation from experimental data. In the case of stable, unreactive ligands, such as inhibitors or allosteric regulators, the complexes can be generated by

co-crystallization or by soaking the ligand into fully grown crystals. Protein complexes with reactive short-lived species that occur in chemical or binding reactions can be determined using monochromatic X-ray diffraction techniques via kinetic trapping approaches. Crystals of macromolecules are characterized by a high content of solvent (typically 30 to 80%), which is arranged in large, interstitial channels spanning the crystals. Affinities may change at crystallization conditions so affinities should be determined under near crystallization conditions[42]. If the protein–ligand affinity is not very high (lower than micromolar) the ligand should be added with the appropriate concentration to all solutions to which the crystal is exposed, particularly the cryoprotectant. As a rule of thumb, to ensure very high occupancy at the binding site of a protein P, the concentration of the ligand L should be about ten times higher than the dissociation constant K_D[43].

The major limitation of this technique is that it requires regular crystals. Crystal production can be difficult and time consuming and sometimes impossible. Due to crystallisation conditions the structure may not wholly represent as it exists in solution. Surface residues may be influenced by crystal packing. Because the surfaces of proteins are usually where they interact with other proteins or with their ligands, these differences, although small, are likely to be biologically relevant.

1.3.3 Fluorescence spectrometry in studies of carbohydrate-protein interactions

Fluorescence techniques are widely used in the study of protein–ligand interactions because of their inherent sensitivity, and the fact that they can be implemented at true equilibrium conditions. Techniques for assessing the affinity used are measurement on fluorescence intensity change, lifetime, polarization anisotropy, and energy transfer[44]. In many cases, protein–ligand interactions can be studied by following changes in fluorescence intensities of a probe or in the intrinsic fluorescence of the protein subsequent to ligand binding. In systems, in which intensity changes do not occur or are minimal, polarization or anisotropy determinations are extremely useful since they rely upon changes in rotational mobility between free and bound probes.

1.3.3.1 Polarisation and anistropy

A small molecule moves and rotates freely and very rapidly in solution, so that a fluorescent probe attached to it will be free of directionality (i.e., isotropic). However, when such a molecule is bound to a large molecule which is relatively slow in rotation and tumbling, then the fluorescent probe on the small molecule can acquire directionality *(i.e.,* anisotropic) and is said to be "polarized." Polarization *(P)* is defined as

$$P= (I_{\parallel}-I_{\perp})/ (I_{\parallel}+I_{\perp})$$

where $I_{\parallel}$ and $I_{\perp}$ are the intensities observed parallel and perpendicular, respectively, to an arbitrary axis. The extent of such polarization is related to tightness of the binding, and therefore measurement of polarization anisotropy can lead to determination of association constant. The most critical consideration is probably the choice of fluorophore. Another critical consideration is the fluorescence lifetime of the fluorophore.

1.3.3.2 FRET

FRET is the physical process by which energy is transferred non-radiatively from an excited molecular chromophore (the donor, D) to another chromophore (the acceptor, A) by means of intermolecular long-range dipole-dipole coupling (0.5-10 nm) and is thus complementary to NOEs, which allow measuring distances in the range of 0.1-0.5nm). For FRET to occur efficiently, the donor emission peak and the acceptor excitation peak must overlap as much as possible. FRET is governed by Förster's equation as shown below:

$$E = R5/(R^5+r^6)$$

where *E* is the efficiency of energy transfer, *r* is the distance between the donor and the acceptor, and *R* is the Forster constant which has to be determined for each specific pair of fluorescent donor-acceptor and the specific experimental conditions[45].

1.3.4 Calorimetric experiments

1.3.4.1 Isothermal titration calorimetry (ITC)

Isothermal titration calorimetry (ITC) is primarily employed to investigate the thermodynamics of ligand-macromolecule interactions. ITC directly estimates the binding thermodynamics through measurement of the energetics of molecular

interactions at constant temperature. A series of data points representing the amount of heat released (exothermic) or absorbed (endothermic) per mole of injectant (usually the ligand) after each injection is plotted as a function of the molar ratio [LT]/[MT] of the total ligand concentration [LT], and the total macromolecule concentration, [MT], to generate the binding isotherm. An example of binding isotherm of concavalin A is shown in the Figure 4.

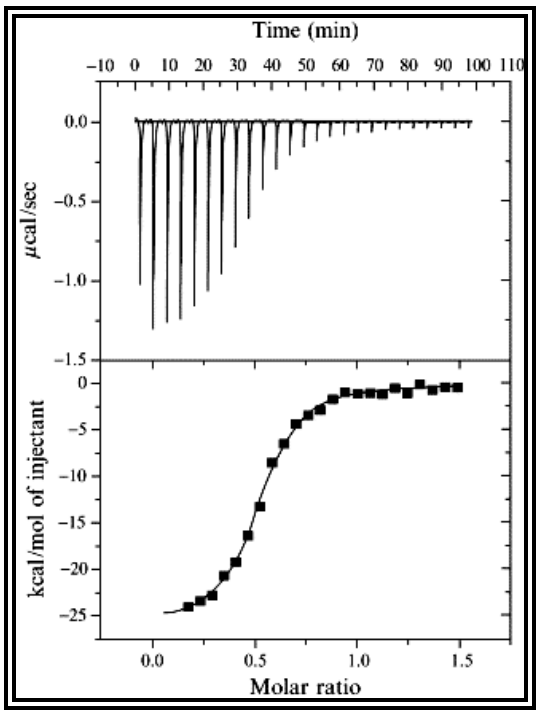


Figure 4: Isothermal titration microcalorimetry (ITC) profile of ConA (0.020 m*M*) with α-d-mannopyranoside bivalent analog 5 substrate (0.39 m*M*) at 27°C. *Top:* Data obtained for 30 automatic injections, 4 µl each, of **5**. *Bottom:* Integrated curve showing experimental points (▪) and the best fit (—). The buffer was 0.1 *M* sodium acetate at pH 5.2 with 0.1 *M* NaCl and $CaCl_2$ and $MnCl_2$ (5 m*M* each)[46].

ITC measurements provide direct determinations of the binding enthalpy ΔH, the association constant, K_a and the number of binding sites (n) of the protein. From measurements of K_a, the free energy of binding, ΔG, can be calculated. The entropy of binding ΔS is obtained from ΔH and ΔG. Thus, ITC measurements can determine the complete thermodynamics of binding of a carbohydrate to a protein. In addition to determining thermodynamic binding parameters, raw ITC data can be used to construct Scatchard and Hill plots to determine whether cooperativity effects occur on binding, which can provide some important aspects of the binding mechanisms of these molecules[46].

1.3.4.2 Differential scanning calorimetry (DSC)

Differential scanning calorimetry (DSC) is an experimental technique to measure the heat energy uptake that takes place in a sample during controlled increase or decrease in temperature. It is used to the study of macromolecular conformation and interactions in solutions at reasonable concentrations and provides some idea how thermal transitions might be effected by ligand binding. ITC is the more direct method to study the protein ligand interactions and DSC can provide some preliminary information about the binding process. The advantages of calorimetric techniques are that they are usually noninvasive and require no chemical modifications or immobilization of any of the reactants. If required the formed complexes could be further studied by other methods. Furthermore with careful analysis and interpretation they directly provide fundamental thermodynamic information about the processes involved.

1.3.5 Surface plasmon resonance

Surface plasmon resonance (SPR) biosensors belong to label-free optical biosensing technologies. Their ability to monitor the interaction between a molecule immobilized on the surface of the sensor and the interacting molecular partner in a solution have made SPR sensors a very powerful tool for biomolecular interaction analysis[47]. The specificity, the strength, the kinetics and some thermodynamic parameters of sugar-protein interactions are easily assessed by surface plasmon resonance (SPR). SPR is particularly suitable to 1) evaluate the macromolecule by checking its ability to bind to its ligand and thus assessing its structural integrity, and 2) for equilibrium analysis (affinity and enthalpy) of weak interactions (K_D >100 μm)[48] and the precise temperature control makes it possible to estimate the binding enthalpy by van't Hoff analysis[49], 3) the real time binding data is used for the analysis of binding kinetics.

Surface plasmon resonance (SPR) imaging is an optical technique that is used to spatially monitor localized differences in the reflectivity of incident light from a prism-gold film interface that result from molecules adsorbing to or desorbing from the gold film. In order to detect an interaction one molecule (the ligand) is immobilized on the sensor surface and the analyte is injected continuously through the flow cell and the

change in refractive index is measured in real time and the result plotted as response or resonance units (RUs) versus time[50] (Figure5). One RU represents the binding of approximately 1 pg protein /mm^2. In practice >50 pg/mm^2 of analyte binding is required. Thus the major advantage of the SPR is that very small amounts of the protein are required and binding between the biomolecular recognition element and analyte can be observed directly without the use of radioactive or fluorescent labels.

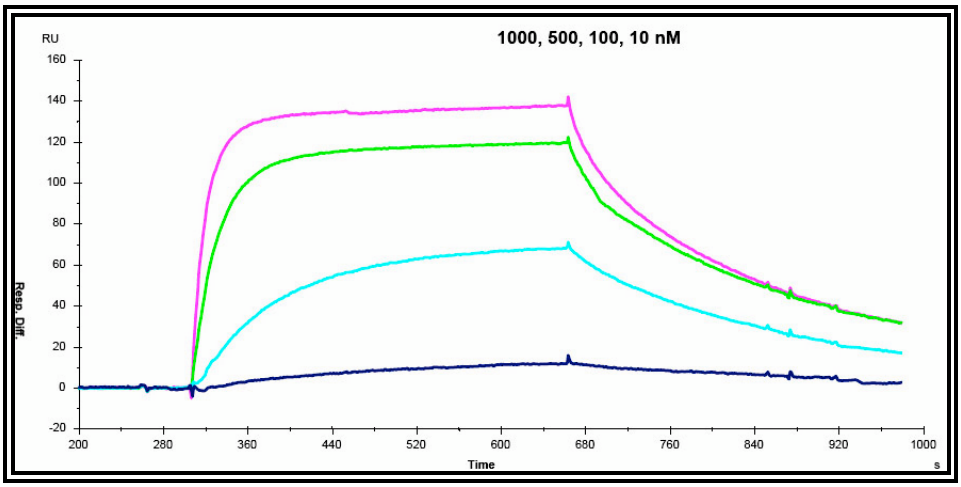


Figure 5: Example of output from Biacore

There are three factors that can influence the interaction between the adsorbing protein and the carbohydrate surface. They are protein-protein interactions between adsorbing protein molecules, the carbohydrate surface site density, and the number of interactions occurring between the adsorbing protein and the carbohydrate surface[51]. Taking care to avoid the pitfalls like aggregation of the analytes, heterogenous immobilization, mass transport limitations etc it is a very useful tool to study protein carbohydrate interactions.

1.4 Molecules under study

1.4.1 Lipopolysaccharides

Approximately some 120 years ago Richard Pfeiffer, a coworker of Robert Koch in Berlin, discovered that cholera bacteria in addition to heat-labile exotoxin produce another toxin[52]. In contrast to the secreted exotoxins this new, heat stable toxin was found to be a constituent of the bacterial cell and, therefore, Pfeiffer termed it endotoxin. Today we know that endotoxin (lipopolysaccharide, LPS) is the main outer

membrane component of gram-negative bacteria and plays a key role during severe Gram-negative infection, sepsis, and shock[53]. The septic shock syndrome caused by endotoxin still has an unacceptably high mortality rate and, owing to increasing numbers of resistant strains, remains an ongoing threat throughout the world.

The presence of low amounts of LPS, are rather beneficial for the host and lead to enhanced resistance to infections and malignancy due to its immunostimulatory activity[54]. However, this picture changes dramatically when larger amounts of LPS are present in the bloodstream, as observed during severe Gram-negative bacterial infections (notably after application of antibiotics) or possibly caused by translocation of enterobacteria from the gut. Released LPS causes various pathophysiological reactions including fever, leukopenia, tachycardia, tachypnoe, hypotension, disseminated intravascular coagulation, and multiorgan failure[55].

1.4.2 Structure and different types of LPS

Lipopolysaccharides are the characteristic components of the cell wall of Gram-negative bacteria. They contribute to the integrity of the outer membrane, and protect the cell against the action of bile salts and lipophilic antibiotics. Lipopolysaccharides are made up of a hydrophobic glycolipid (lipid A, which is responsible for the toxic properties of the molecule), a hydrophilic core polysaccharide chain, and a hydrophilic O-antigenic polysaccharide side chain (Figure 6). If a wild strain of bacterium is irradiated with UV light or exposed to mutagenic compounds, it will mutate and LPS with shorter polysaccharide chains Ra, Rb, Rc, Rd, Re, etc may be formed where a, b, c, etc. designate 1st, 2nd, 3rd, etc. degree of the polysaccharide length of a given LPS (Figure7). Ra and Re designate the mutants with the longest and shortest chain lengths, respectively. The most extreme mutants are the Re mutants which produce an LPS which is made up of lipid A and 3-deoxy-D-manno-octulosonic acid (2-Keto-3-deoxyoctonate, KDO) as the sole constituent of the core[56]. Removal by hydrolysis of the polysaccharide chain from LPS produces lipid A, either as the naturally occurring, cytotoxic diphosphoryl form[57] or the less toxic monophosphoryl form[58].

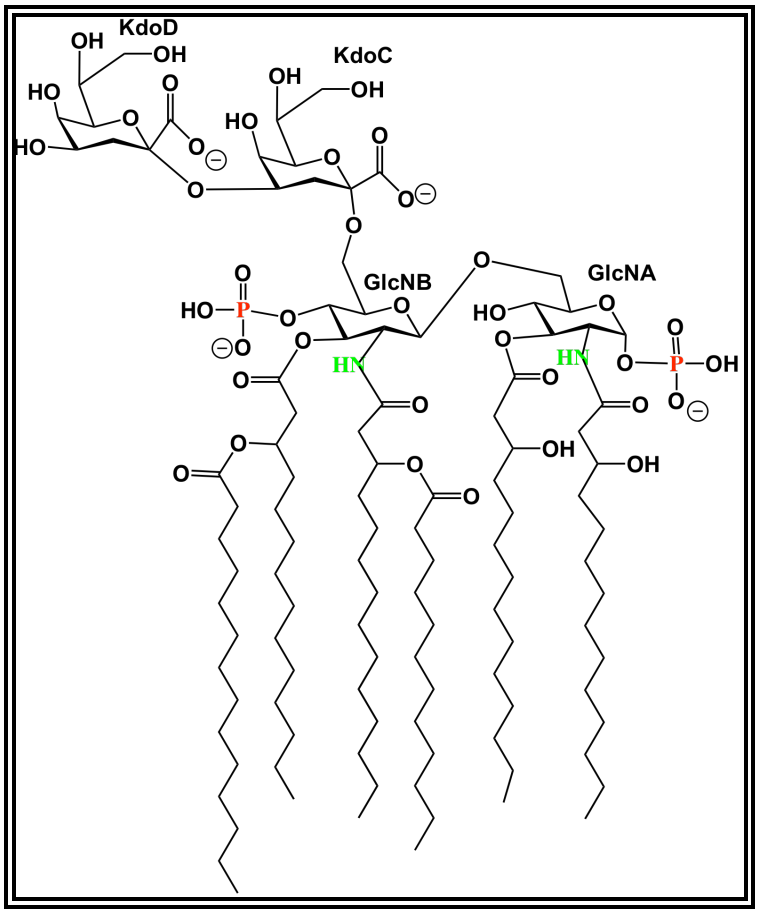


Figure 6: Chemical structure of E. coli Re LPS.
Kdo, 2-Keto-3-deoxyoctonate; GlcN, N-acetylglucosamine.

LPS was shown to occur in three different structures in the outer membrane: in a lamellar orientation, as a hemi-micelle complexed with proteins and as hemi-micelles introduced by divalent cations and/or polyamines[59]**.** Neutralisation in part of the negative charges of LPS by metal cations helps to stabilize the membrane by decreasing the strong electrostatic repulsion between the highly negatively charged LPS molecules. Ca^{2+} and Mg^{2+} ions are primarily essential for the stability of the membrane. The phosphoryl groups on the LPS as well as the carboxyl group on one of the KDO units were shown to be involved in binding $Ca^{2+}$and Mg^{2+}. This was confirmed by metal binding studies conducted with heptoseless mutants of *E.coli* by ^{13}C and ^{31}P nuclear magnetic resonance[60].

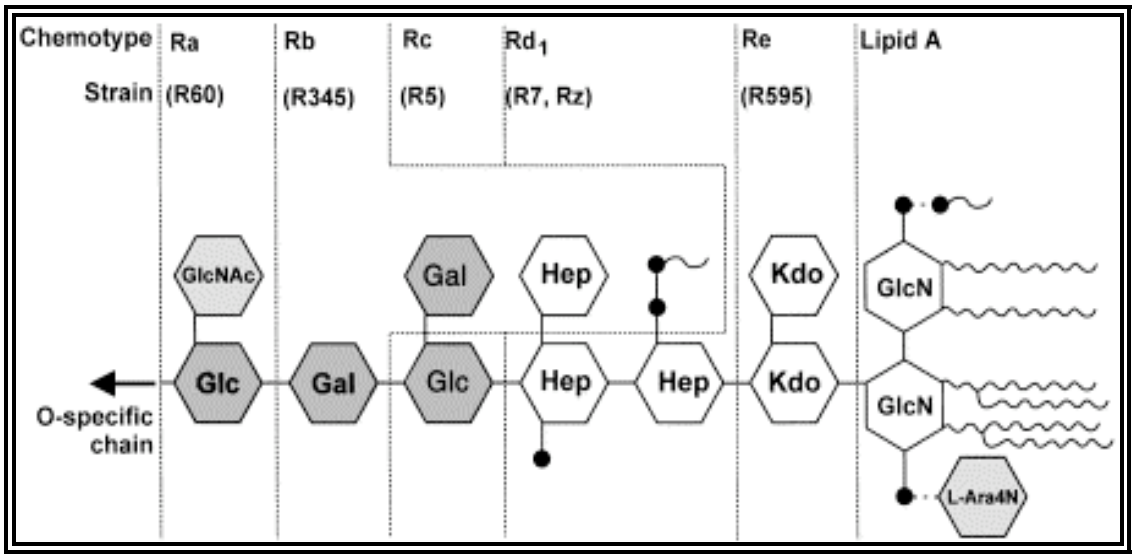


Figure 7: Schematic structure of various LPS rough mutant LPS from S. minnesota. The O-specific polysaccharide consists of a varying number of repeating units of a pentasaccharide. Abbreviations: Hep, L-glycero-D-manno-heptose; Gal, galactose; Glc, glucose; GlcN, glucosamine; GlcNAc,N-acetylglucosamine; Kdo,3-deoxy-D-manno-oct-2-ulosonic acid; L-Ara4N, 4-amino-4-deoxy-L-arabinose; (•) phosphate. Beside the two phosphate groups in the lipid A backbone, further phosphate or pyrophosphoethanolamine groups are present in the heptose region for the strains Rz, R5, R345, and R60; (– – –) indicates non-stoichiometric substitution.

Removal of divalent cations by chelators like EDTA results in increased outer membrane permeability, structural perturbations and at higher concentrations extraction of LPS and/or LPS-protein complexes from the cell surface. Similar effects may be observed with various polycations including polymyxins, aminoglucosides etc. Lipopolysaccharide (LPS) and in particular the core domain lipid A have been recognized as valuable targets for antimicrobial agents, e.g. stretogramins, glycopeptides, oxazolidines, glycylcyclines, ketolides, β-lactams, β -lactamase inhibitors and cationic peptides[61]. Furthermore, a larger number of LPS-binding peptides were discovered, spanning from α-helical peptides (e.g. mellitin, efrapeptin) or β-sheet forming peptides (e.g. gramicidin S, tyrocidin A). Very few investigations have been published, in which interactions of Magainin2, an antimicrobial peptide, with LPS incorporated into liposomes were determined; however, no data at atomic resolution were presented[62].

Structural information about the conformation of LPS during immune recognition and signaling events is very important to design ligands capable of neutralizing LPS. Until recently the only 3D structure of an intact LPS molecule that is available was obtained by X-ray crystallography in an incidental complex with the *E. coli*

ferrichrome-iron receptor FhuA[63]. Very recently in combination with isotope labeling by ^{13}C and ^{15}N, assignment of LPS in DHPC micelles was achieved by a combination of a large number of modern heteronuclear solution NMR techniques.

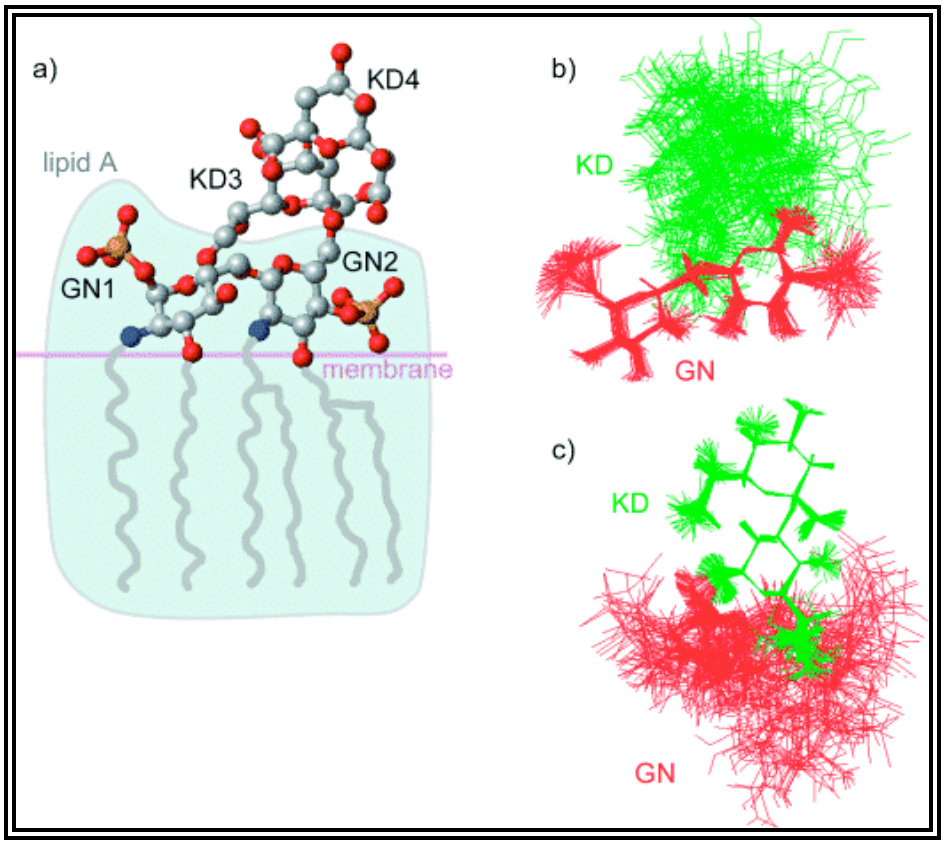


Figure 8: a) The lowest energy structure of LPS-6 with the GN (GlcN) and KD (KDO) units shown in a ball-and-stick representation and the fatty acid chains shown as a cartoon. The lipid A part of LPS-6, which is responsible for the endotoxic reaction, is highlighted by a pale blue background. The orientation of LPS with respect to the membrane is indicated. b, c) Calculated structure ensemble showing the 50 lowest energy structures with the units of GN (b) or KD (c) aligned relative to each other[64].

The structure (Figure 8) indicates that all GlcN and KDO units of LPS-6 (LPS in which the fatty acid chains include only atoms up to C4 atom) are in the glucopyranose configuration with a corresponding chair sugar pucker. GlcN-1 is in the α-d configuration ($^3J_{H1H2}$ = 2.7 Hz), whereas GlcN-2 is in the β($^3J_{H1H2}$ = 9.6 Hz) configuration. Based on sugar-sugar and sugar–fatty acid NOEs, it was found that GlcN1 shows numerous contacts to Fatty acid 3 (FA3) which is consistent with a tilted conformation of GlcN-1 relative to GlcN-2. Numerous strong intra-subunit NOEs

between the KDO units indicates restricted rotation around the KDO-KDO glycosidic linkage. The relative orientation of the GlcN and KDO units is not well defined due to the intrinsic dynamics of the molecule[64].

1.4.3 Structural requirements for the bioactivity of LPS

Although there is a great compositional variation among endotoxins derived from different bacterial serotypes, they all share a common structural principle. The minimal requirement for lipidA bioactivity, referred to as the cytokine inducing capacity, is a molecule having two pyranoglucose-configurated hexosamine residues, two phosphoryl groups, and six fatty acids as present in *E. coli* lipidA. LipidA structural differences-such as hydroxylation, secondary substitution, as well as the presence of longer-than-average length fatty acid chains (C16)-might be potentially regulated by the presence of specific environmental factors (pH, divalent cations) and specific regulatory mechanisms that respond to these factors[65].

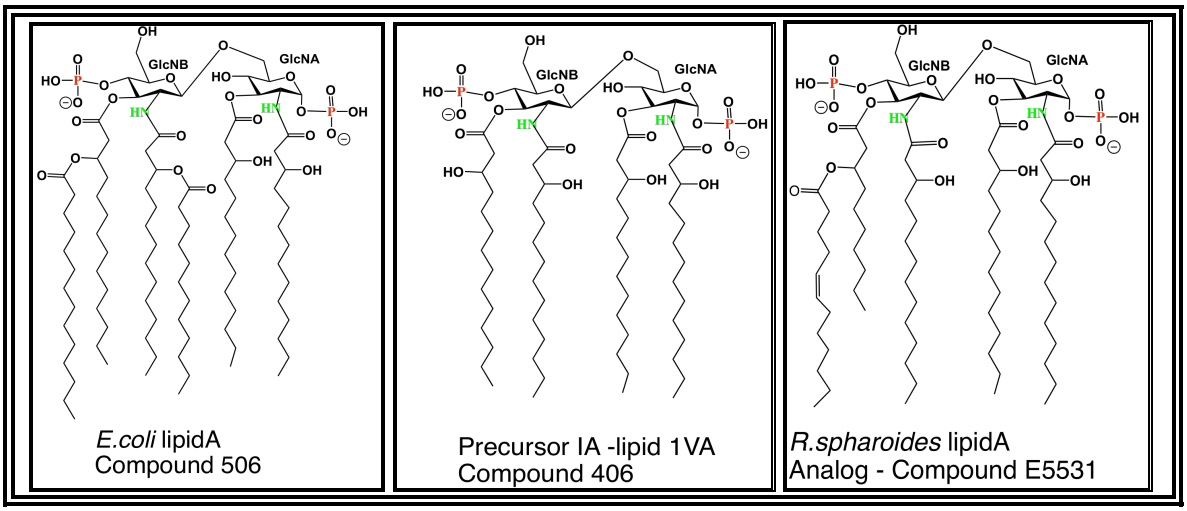


Figure 9. Chemical structure of endotoxin, lipid A and lipid A analogs

LipidA partial structures deficient in one of these elements are less active or non-active regarding the induction of monokines in human monocytes. The synthetic lipidA partial structures with the 1'-dephospho derivative and the 4'-dephospho derivative were found to be less active than the bisphosphorylated compound, which signifies the importance of the two-phosphoryl groups for bioactivity. Also, the presence as well as the position of the six fatty acids is of critical importance for the full bioactivity of LPS. The lack of the two secondary fatty acids as in the tetraacylated lipid A precursor Ia (also known as lipid IVa or LA-14-PP) **(**Figure 9**)** makes this

molecule completely inactive[66, 67]. Additional acylation as in the highly acylated heptaacyl lipid A (*S. Minnesota* lipid A, compound 516) leads to less bioactivity[68]. A potent synthetic LPS-antagonist *in vitro* as well as *in vivo* is compound E5531 (Figure 10) which was synthesized based on the structure of the *Rhodobacter spharoides* lipid A[69].

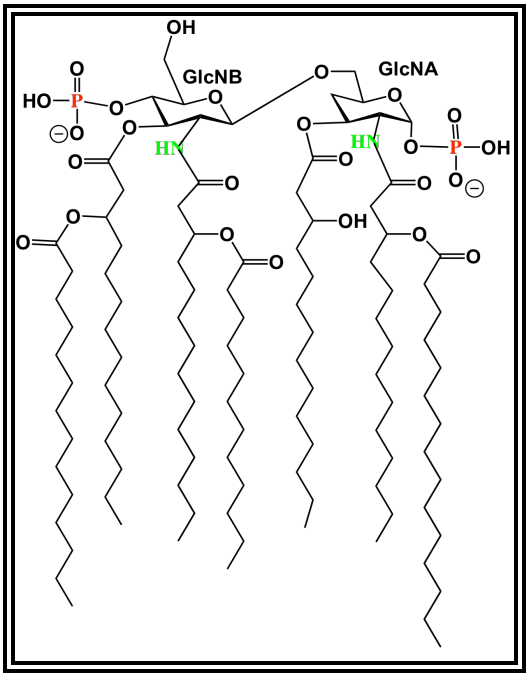


Figure 10: Synthetic LPS-antagonist compound E5531

In addition to the requirements on the chemical structure of LPS or lipid A also the physical structure of LPS has an influence on its bioactivity. LPS or lipidA, as amphiphilic substances form aggregates in aqueous solution with a distinct three-dimensional structure. The bioactive lipid A was shown to form nonlammellar structures, which are either cubic (*S. Minnesota*) or hexagonal (*R. gelatinosus*)[70], whereas the inactive lipid A from *R. capsulatus* forms lamellar aggregates. However, this conformational difference between the compounds may only be of significance at higher LPS/lipid A concentrations, since it has been found that endotoxicity is expressed by monomeric LPS molecules[71].

1.4.4 Biosynthesis of LPS

LPS is in general are synthesized as two separate components, the lipid A core and the O polysaccharide, which were then ligated to give the complete molecule.

Mutants, which make incomplete LPS, are known as rough mutants because the colonies on agar plates are flatter, spread further and lack the normal smooth appearance of wild type of colonies. LipidA is synthesized on the cytoplasmic leaflet of the inner membrane by several steps of acylation of UDP GlcNac with the help of acyl carrier protein (ACP) and various specific acyl transferases. The core sugar residues are sequentially added on it by specific glycosyltransferases, and the completed lipidA core is translocated to face the periplasmic side of the inner membrane.

The heteropolymeric O-units are synthesized on a lipid carrier molecule, undecaprenyl phosphate (Und-P). A complete O-unit is flipped to the periplasmic face of the inner membrane where Wzy (the O-ag polymerase) polymerizes the O-units and Wzz (the O-ag chain length determinator) controls the length of the O-antigen by a not yet understandable mechanism. The homopolysaccharide O-chains synthesized onto Und-P are completed to their full lengths in the cytoplasm after which they are translocated to the periplasmic space by the Wzm and Wzt proteins that are components of an ATP binding cassette transporter system. The ^{13}C-enrichment of LPS is valuable since NMR assignments could be obtained on the basis of ^{13}C-^{13}C connectivities in uniformly labeled *E.coli* LPS residues. This can be achieved by addition of labelled D-[1-^{13}C] glucose to defined M9 medium, which directly incorporates glucose as the building block for the hexose skeletons in the polysaccharide. Biosynthesis of acyl substituents occurrs via the triose pool followed by decarboxylation to give acetyl building blocks labeled with ^{13}C at the methyl group. ^{15}N labeling of amino groups of the glucosamine moiety is achieved by adding ^{15}N NH_4Cl to the M9 medium. Since LPS is a component of Gram-negative bacterial outer membranes and established protocols for isolating LPS are available, isotopically enriched LPS can be obtained in a relatively easy and economical manner.

1.4.5 LPS signaling pathways

LPS activates monocytes and macrophages to produce cytokines such as TNF-α, IL-1, and IL-6, which, in turn, serve as endogenous mediators of inflammation through receptor-mediated interactions with various target cells. This phenomenon is known as "endotoxin tolerance," while excessive activation of monocytes and macrophages by a bolus of LPS can lead to septic shock, a serious systemic disorder that can result in

multiple organ failure and death. The molecular mechanisms underlying these extremely different responses are as follows. LPS is disaggregated into monomers by LPS-binding protein and presented to cell-surface CD14. Toll-like receptors (TLR4) then bind to the CD14–LPS complex and gets activated, stimulating a cytoplasmic signal transduction cascade (Figure 10), which leads to liberation of free nuclear factor-κB (NF-κB) and increased transcription of inflammatory genes including tumor necrosis factor-α (TNF-α), interleukin-1β (IL-β), and inducible nitric oxide synthase (iNOS). The CD11/CD18 integrins are transmembrane proteins. Like CD14, they are also capable of mediating LPS-induced cellular activation[72]. LPS can also be internalized through an yet unknown mechanism, and may bind to intracellular Nod1 and result in either activation of transcription factors or processing of a hypothetical protein X, which may serve as a true ligand for TLR4.

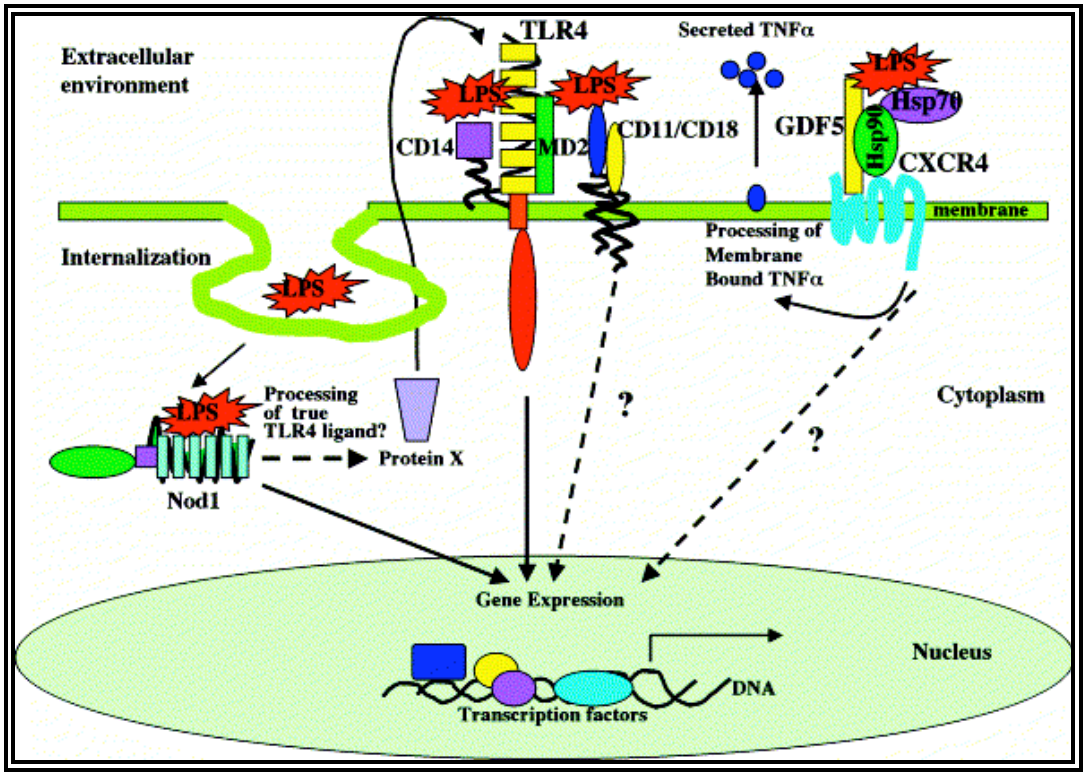


Figure 10: Schematic representation of LPS signaling[73].

Although TLR4 receptors are presumably the main LPS signaling receptors they are not solely responsible for LPS recognition. After the ligation of CD14 by LPS, different signalling molecules must be recruited to the site of ligation, where LPS is then briefly released into the lipid bilayer. There it interacts with a complex of receptors, which includes Hsp70, Hsp90, CXCR4, GDF5 and TLR4[74]. Thus, it is

likely that the assembly of multiple membrane molecules form a LPS receptor domain (LRD) occurs. It is well known, that GPI-linked molecules accumulate together with cholesterol and glycosphingolipids forming special membrane microdomains called *rafts* or "detergent insoluble glycolipid-enriched domains" (DIG)[75-77]**.** These domains are the loci of numerous cell functions from membrane traffic and cell morphogenesis to cell signaling. LPS has been shown to be able to activate the NF-kB, ERK1/2 and SAPK/JNK pathways. It is possible that different cell types and different supramolecular complexes utilize different intracellular pathways[74].

1.4.6 Key players of LPS signaling pathway - Endotoxin-binding proteins

1.4.6.1 LBP

LBP is a soluble acute phase protein that binds to bacterial lipopolysaccharide to elicit immune responses by presenting the LPS to important cell surface pattern recognition receptors called CD14 and TLR4. LBP is synthesized by hepatocytes and intestinal epithelial cells and is present in normal serum at concentrations of 5 to 10 μg/ml, rising up to 200 μg/ml 24 h after induction of an acute-phase response triggered by interleukin-1 (IL-1). LBP has a concentration-dependent dual role: low concentrations of LBP enhance the LPS-induced activation of mononuclear cells (MNC), whereas the acute-phase rise in LBP concentrations inhibits LPS-induced cellular stimulation. LBP binds a variety of LPS (endotoxin) chemotypes from rough and smooth strains of gram-negative bacteria and even lipid A, the lipid moiety of LPS[78].

1.4.6.2 BPI

Bacterial/permeability-increasing protein is a 50 kDa protein first isolated from azurophilic granules of neutrophils by Weiss and colleagues[79]. BPI is released from neutrophils after stimulation by a variety of compounds including LPS, *N*-methyl-formylated chemotactic tripeptides (e.g., f-Met-Leu-Phe), and phorbol esters. It is found in extremely high concentration in abscess fluid. BPI has LPS-binding and bactericidal activity specific for Gram-negative bacteria and it binds avidly to LPS in a lipid A-dependent manner. However, BPI lacks the activity of LBP that mediates LPS binding

to CD14 and to HDL. The amino-terminal region of BPI, ($BPI_{1–197}$, rBPI23) retains full LPS-binding activity[80].

1.4.6.3 CD14

CD14 is a GPI-anchored protein constitutively expressed on the surface of various cells, including monocytes, macrophages, polymorphonuclear neutrophils, chondrocytes, B cells, dendritic cells, gingival fibroblasts, keratinocytes, and human intestinal epithelial cell lines. Aside from this membrane-bound (mCD14) state, a truncated form of CD14 is also found in a circulating soluble (sCD14) form in serum at a concentration of 5-6 μg/ml. Both mCD14 and sCD14 causes LPS mediated activation of cells. CD14 has been shown to sensitize the cell types that do not express membrane-bound CD14, such as endothelial cells, astrocytes and epithelial cells. It has recently been demonstrated that sCD14 is a regulatory factor capable of modulating cellular and humoral immune responses by interacting directly with T and B cells. Mutant sCD14 containing the N-terminal 152 amino acid residues were found to have activity equivalent to the full length sCD14[81]. Mutational studies indicate that LPS binding and cellular activation are separate functional properties of CD14[82]. Studies with CD14 derived peptides showed that the peptide comprising amino acids 81 to 100 is sufficient to bind and neutralize LPS[83]. The binding of LPS to CD14 can also occur in the absence of LBP but only in the presence of a high concentration of LPS; however, LBP markedly accelerates this activity.

Besides binding LPS, CD14 may function as a receptor for peptidoglycan, the major cell wall component of Gram-positive bacteria[84], other microbial products with similar structural features[85], human heat shock protein 60[86], and other ligands (ceramide, anionic phospholipids, modified lipoproteins, opsonized particles)[87]. Thus CD14 can be defined as a pattern recognition molecule in innate immunity. It also has been found that the interaction of LPS/LBP with CD14 causes the exchange of LPS with lipids in target membranes[87-89], and this lipid transfer may be responsible for LPS-induced signal transduction. The rate of the exchange reaction depends on the lipid composition of the target membranes, which has led to the speculation that CD14 functions only to direct LPS insertion into particular membrane domains[81]. High concentrations of sCD14 (10–30 $\mu g/mL^{-1}$) could compete with mCD14 on myeloid

cells for LPS binding and thereby neutralize partly the biological effects of the endotoxin molecules[90]. This LPS neutralizing property of sCD14 makes it a potential therapeutic agent in septic patients. Thus the exact role of CD14 in physiological and pathological situations is not well defined. CD14 was initially described as a specific receptor for LPS. At the monocyte surface, the LPS receptor mCD14 and the LPS-binding protein (LBP) interact with LPS, and form a high affinity trimolecular complex that allows monocytes to detect the presence of LPS. In contrast, two opposite functions have been described for sCD14. It can either reduce endotoxin-induced activities by competing with mCD14 for LPS binding, or it mediates the LPS-induced activation of non-CD14-expressing endothelial, epithelial, and smooth-muscle cells.

1.4.6.4 Structure of CD14

The monomeric subunit of CD14 contains thirteen β strands, and 11 of them, from β3 to β13, overlap with conserved leucine-rich repeats (LRRs) (Figure 11). The concave surface of the horseshoe shaped structure consists of a large β-sheets of 11 parallel and two antiparallel β-strands. The convex surface of CD14 contains both helices and loops, in no regular pattern. A large hydrophobic pocket was found on the NH_2-terminal side. The pocket is the main component of the lipopolysaccharide-binding site. Dimerization in the crystal is mediated by residues in β 13 and in the loop between β12 and β13. Parallel β-sheets from the two monomers interact in an antiparallel fashion and form a large and continuous β-sheet encompassing the entire CD14 dimer[91]. Unlike mammalian CD14, chicken CD14 appears to have trans-membrane and cytoplasmic domains, suggesting that it is not GPI-anchored. This has important implications for the role of CD14 in the chicken's response to LPS. Reports suggest that chicken cells respond less well to LPS than the equivalent mammalian cells[92].

Figure 11: Structure of the CD14 dimer. Two monomers of CD14 in the crystal are colored in *green* and yellow. Disulfide bridges are shown as red spheres. mCD14 is GPI anchored membrane protein. Soluble sCD14 lacks the GPI anchor and mutant sCD14 containing N-terminal 152 amino acid residues sCD14 were found to have activity equivalent to the full length sCD14 and this construct was used for our studies.

1.4.6.5 Toll-like receptors

Although there is no doubt that CD14 binds LPS, CD14 is not capable of initiating a transmembrane activation signal because it is a glycosyl phosphatidyl inositol (GPI)-anchored protein and has no transmembrane domain. The signaling protein that mediates this phenomenon is likely to be a member of the Toll-like receptor (TLR) family. TLRs are a type of *pattern recognition receptor* (PRR) that are now counted among the key molecules that alert the immune system to the presence of microbial infections. They are named for their similarity to Toll, a receptor first identified in the fruit fly *Drosophila melanogaster*, and found to have an essential role in the fly's immunity to fungal infection[93], which is achieved by activating the synthesis of antimicrobial peptides. Members of the Toll-like receptor (TLR) family recognize conserved microbial structures, such as bacterial lipopolysaccharide and viral

double-stranded RNA, and activate signaling pathways that result in immune responses against microbial infections. All TLRs induce stereotyped responses such as inflammation. However, individual TLRs can also induce immune responses that are tailored to a given microbial infection. Thus, these receptors are involved in both innate and adaptive immune responses[94].

1.4.7 LPS responsive cells

1.4.7.1 Monocytes

The most prominent LPS-sensitive cell populations are cells of the monocyte/macrophages lineage. These cells produce a large variety of bioactive protein mediators in response to LPS including interleukin 1 (IL-1), IL-6, IL-8, and in particular tumor necrosis factor α (TNF-α)[54]. A lot of host cells respond to these cytokines and thereby initiating the typical acute phase response (which helps to eliminate the invading microorganisms. The massive release of cytokines, however, becomes hazardous for the organism by causing shock, cell damage, and multi-organ failure[95]. In addition to cytokines, macrophages produce reactive oxygen species (superoxide anion, hydrogen peroxide, hydroxyl radicals and nitric oxide) on exposure to endotoxin, which also been shown to be involved in the pathophysiology of septicemia.

1.4.7.2 Polymorphonuclear leukocytes (PMN)

The phagocytosis of microorganisms by PMN is drastically enhanced in the presence of LPS and presents a good example for the nonspecific immunostimulatory capacity of endotoxin. PMN are able to neutralize LPS since they contain enzymes that are able to degrade (deacylate and dephosphorylate) LPS and lipid A to nontoxic partial structure derivatives[96]. Furthermore, they express and release antimicrobial proteins with strong affinity to LPS and thus neutralize LPS bioactivity. One of these proteins is the cationic 55 kDa bactericidal/permeability-increasing protein (BPI), which is able to kill bacteria by binding to the bacterial surface.

1.4.7.3 B- and T-lymphocytes

The polyclonal activation of murine B lymphocytes in response to LPS, resulting in proliferation, differentiation, and the secretion of immunoglobulins, seems to be an early defense mechanism, since it leads to the enhanced release of antibodies with various antimicrobial specificity[97]**.** Activation of T-lymphocytes by LPS depends on accessory monocytes providing costimulatory signals[98]. T-lymphocytes can be downregulated by lipid A leading to an increased antipolysaccharide antibody production[99]. All these findings support the hypothesis that LPS indeed is involved in cellular immunity against microorganisms.

1.4.7.4 Vascular cells, epithelial cells

Vascular cells (endothelial cells and smooth muscle cells) as well as epithelial cells can respond to LPS by the release of cytokines, for example IL-1, IL-6, and/or IL-8[100]. Additionally, they can produce several other mediators such as prostacyclin, nitric oxide, platelet-activating factor, interferons, and colony-stimulating factors.

1.4.8 Strategies of endotoxin antagonism

Many steps along the pathway from the interaction of LPS with the mammalian cell membrane to intracellular signaling events have been considered as potential drug targets. They are (1) identification of naturally occurring lipopolysaccharide (LPS)-binding proteins; (2) generation of novel LPS-binding antibodies, proteins and peptides; and (3) characterization of the molecular determinants of LPS binding and making LPS antagonists by modification of lipid A structure[101].

1.4.8.1 Anti-endotoxin antibodies

Polyclonal or monoclonal antibodies to the outer ‘O’ antigens of LPS provide serotype specific protection in experimental Gram-negative infection and in patients, which offers too limited range of protection against which is against only specific strains of bacteria[102]. Antibodies against deep core lipid A region (Anti-DCLA mAbs) demonstrate cross-reactive binding activity and protect against many strains of bacteria but binding of these agents to smooth bacteria and its LPS have been inconsistent and these agents appear to be markedly less potent than those directed against the O-antigen polysaccharide region[101].

1.4.8.2 LPS-binding proteins

The identification of LPS-binding proteins such as LBP, sCD14, BPI, and LALF has stimulated the development of these anti-LPS agents for preclinical and clinical evaluation. LALF is the most avid LPS-binding agent yet identified *in vitro*[103]. The amino-terminal domain of BPI ($BPI_{1–197}$) contains the full LPS neutralizing and bactericidal activity of the parent molecule. LBP-LPS complexes could be neutralized by soluble CD14, which would prevent their interaction with macrophages and the induction of TNF-α release. Thus soluble CD14 could be a natural inhibitor of the deleterious effects of endotoxin[104]. As discussed above, LBP, CD14, and TLR(s) sequentially bind LPS, ultimately leading to transduction of a pro-inflammatory signal in LPS-responsive cells. The blockade of each successive step in this chain has been considered as a potential site of therapeutic intervention.

1.4.8.3 Anti-LPS peptides

One reason that anti-DCLA mAbs and BPI bind inconsistently to smooth strains of bacteria or their derived LPS may be that the DCLA region of LPS is relatively inaccessible, being embedded in the outer membrane and shielded by external portions of the LPS molecule. The smaller peptide derivatives should more readily access lipid A, if steric factors significantly hinder the binding of larger proteins. Endogenous host defense peptides, fragments of endotoxin-binding proteins and synthetic anti-endotoxin peptides fold into α-helical, β-hairpin, extended and compact conformations without regular secondary structure. In animal models many of the peptides have demonstrated good *in vitro* and *in vivo* endotoxin-neutralizing activity, but up to now none of the peptides has been approved for clinical application with an anti-endotoxin indication[105].

1.4.8.4 Antimicrobial peptides

Antimicrobial peptides (AMPs) are crucial humoral components of the innate immunity system of virtually all organisms, by which they defend themselves from the invasion of attacking pathogens[106]. They had already existed for at least 500 million years[107], before the first glimmerings of the adaptive immune appeared. Hundreds of these gene-encoded peptides, usually ranging in size from 12 to 50 residues, have been isolated from bacteria, fungi, plants, and animals, including humans. Despite broad

divergences in sequence and taxonomy, most antibiotic peptides share a common mechanism of action, i.e., membrane permeabilization of the pathogen. Most AMPs are amphipathic and cationic molecules. Their net positive charge promotes their binding to the membranes of microbes, which are generally negatively charged.

Many AMPs act by compromising the structural and/or functional integrity of microbial membranes[108], but some may have additional or alternative modes of action[109]. Factors believed to be important for antimicrobial activity have been identified, including peptide hydrophobicity, the presence of positively charged residues, an amphipathic nature that segregates basic and hydrophobic residues, and secondary structure. High hydrophobicity should force a strong partition of the peptide into the hydrophobic core of the lipid bilayers, regardless of the phospholipid head group, thereby permeabilizing membranes of both eukaryotes and prokaryotes[110]. In addition to their antimicrobial activities, many AMPs also bind strongly to LPS, a property required for those peptides selective for gram-negative bacteria which must interact with the outer membrane before reaching the cytoplasmic membrane and killing the cell[105].

Many of the host defense peptides, which have antimicrobial activity, are also able to bind to and neutralize LPS; however, these two activities do not necessarily correlate. Polymyxin B (PMX-B) is one of the most efficient compounds for the treatment of septic shock, due to its ability to bind and detoxify bacterial lipopolysaccharide. The biophysical properties and the mode of action of AMPs with LPS determine their biological function, the susceptibility of bacteria to them, the ability of LPS to activate the immune system, as well as the potential of AMPs to interfere with this activation and to neutralize LPS-induced endotoxic shock. Better understanding of the peptides characters involved in antibacterial activity, LPS neutralization, toxicity and stability should help us to develop improved antimicrobial peptides with desirable properties[111].

1.4.9 Polymyxins

Polymyxin B is a strongly cationic cyclic polypeptide antibiotic isolated from *Bacillus polymyxa*[112]. Polymyxin B and the other polymyxin antibiotics act primarily by binding membrane phospholipids and disrupting the cytoplasmic membrane,

inducing pore formation in bacterial walls. Polymyxin B binds to the lipidA portion of the lipopolysaccharide in the cell membrane of gram-negative bacteria. It neutralizes the effects of endotoxin at molecular[113, 114] and cellular[115, 116] levels and in animal models of endotoxaemia[117]. The interaction was postulated to involve ionic forces between amino groups in Polymyxin B and phosphate and carboxyl groups in the lipid A-Kdo region, with hydrophobic interactions between the respective acyl groups. Data suggest a stoichiometric binding of one LPS monomer to one polymyxin B molecule[118, 119]. Polymyxin B has a bactericidal action on most gram-negative bacilli, except *proteus spp*. and is not active against Neisseria species, most fungi and Gram-positive bacteria. Activity is inhibited by Fe(II), Co(II), Mn(II) and Mg(II) ions. The use of polymyxins is associated with considerable toxicity, mainly nephrotoxicity and neurotoxicity[120]. But polymyxins have been recently re-introduced in clinical practice for the treatment of patients with multidrug-resistant gram-negative bacterial infections. Data from the recent literature suggest that the use of polymyxins is associated with lower and less severe toxicity compared to that reported in the old literature[121].

1.4.9.1 Structure of polymyxins

The general structure (Figure 12) comprises a cyclic heptapeptide bound to a linear tripeptide, with six unnatural amino acid such as 2,4-diaminobutyric acid (Dab). The γ-amino group of Dab-4 is linked via an amide bond to the C-terminus of Thr-10, while its α-amino group is connected to Dab-3 of the linear tripeptide. The N terminus of Dab-1 is end-capped with a long alkyl chain. They contain one or two D-aminoacids. The presence of positive charges, the cyclic structure, the amphipathicity and the hydrophobic tail are essential for activity against gram-negative microorganisms[122].

1.4.9.2 Types of polymyxins

There are several types of polymyxins, which are produced by fermentation from different strains of *Bacillus polymyxa*. They differ in amino acid composition at position 3, 6, 7, as well as in the nature of the fatty acid (Table 4). We studied the interactions of Polymyxin B, E and M with *E.coli* Re-LPS. Polymyxin M (Figure 13) is biosynthetically produced from the *B.polymyxa* strain.

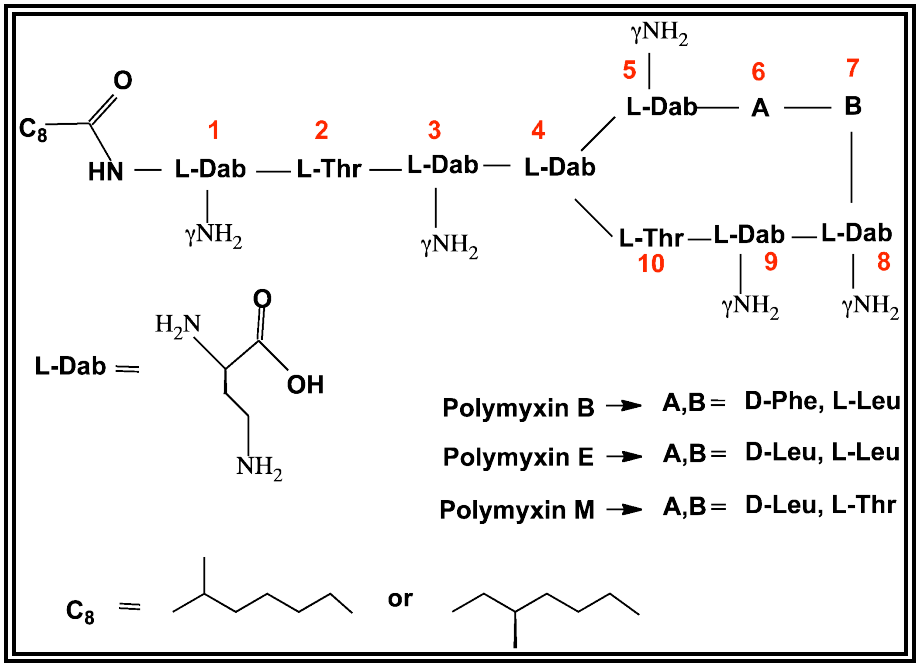


Figure 12: General structure of polymyxins where A can be D-Leu or D-Phe and B can be L-Thr, L-Leu, L-Phe or L-Ile depending on the type of polymyxin.

Type of polymyxin	Composition of variable regions			
	Fatty Acid chain	3	6	7
Polymyxin A	6-Methyl octanoic acid	D-Dab	D-Leu	L-Thr
Polymyxin B1	6-Methyl octanoic acid	L-Dab	D- Phe	L- leu
Ile-B1	6-Methyl octanoic acid	L-Dab	D-Phe	L-Ile
B2	6-Methyl heptanoic acid	L-Dab	D-Phe	L-Leu
B3	Octanoic acid	L-Dab	D-Phe	L-Leu
B4	Heptanoic acid	L-Dab	D-Phe	L-Leu
B5	NA	L-Dab	D-Phe	L-Leu
B6	3-OH 6-Methyl octanoic acid	L-Dab	D-Phe	L-Leu
Polymyxin D 1	6- Methyl octanoic acid	D-Ser	D-Leu	L-Thr
Polymyxin E1	IOA	L- Dab	D-Leu	L-Leu
E2	IOA	L -Dab	D -Leu	L-Leu
Polymyxin M	6-Methyl octanoic acid	L-Dab	D-Leu	L-Thr

Table 4: Composition of variable regions in different types of polymyxins

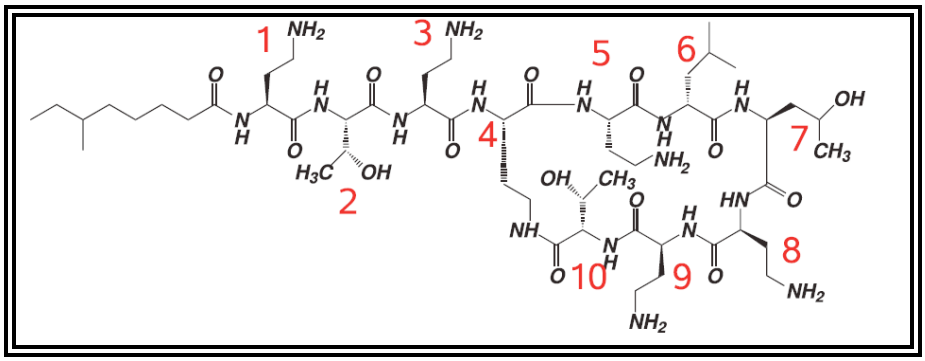


Figure 13: Structure of Polymyxin M

1.4.9.3 Mechanism of action

Polymyxins, which behave like detergents, are rapidly bactericidal. Their target site is the bacterial outer cell membrane. Polymyxins increase permeability of the cell envelope, which lead to leakage of cell contents and, subsequently, cell death. Polymyxins through electrostatic interactions bind to anionic lipopolysaccharide by displacing magnesium (Mg^{2+}) and calcium (Ca^{2+}) from the negatively charged LPS. These cyclic decapeptides are created from nondescript sequences of amino acids lacking unique epitopes that could serve as the recognition site of a protease required for selective destruction of the antibiotic in the presence of cellular protein constituents. Neutralization and sequestration of bacterial lipopolysaccharide, which plays a key role in gram-negative sepsis is required to block the progression of sepsis at early stages in addition to destroying bacteria. Thus, one can envision the use of polymyxin in combination with conventional antibiotics to increase killing and at the same time, neutralize LPS released by these antibiotics.

1.5 Scope of our study

Sepsis is a serious source of mortality in many cases accounting for approximately 200,000 deaths in the US annually[123]. Owing to increasing numbers of resistant strains, it remains an ongoing threat throughout the world. It is known that many antibiotics stimulate the release of endotoxin and thus stimulate the occurrence of such symptoms[124, 125]. Thus, there is great interest in the development of antimicrobial agents, which can reverse rather than promote sepsis, especially given the recent disappointing clinical performance of antiendotoxin therapies.

Our study was devoted to the interaction of lipopolysaccharide from gram-negative bacteria with polymyxins. We present a study of interactions of polymyxin B, E and M with LPS from the deep rough mutant strain of *E.coli* in membrane-mimetic dodecylphosphocholine (DPC) micelles. First a method for efficient purification of biosynthetically produced LPS using RP-HPLC in ternary solvent mixture was developed. Polymyxin M was biosynthetically produced from the *B.polymyxa* strain. Then LPS was incorporated into a DPC micelles and its interaction with polymyxins was studied by heteronuclear NMR spectroscopy. The chemical shift mapping data using isotope labeled LPS or labeled polymyxin as well as isotope-filtered NOESY experiments reveals the mode of interaction LPS with polymyxins. Using simulated annealing protocol and restraints derived from intermolecular NOEs we have obtained the model of LPS – PMX-B complex, which was also in accordance with the chemical shift mapping data. The model was further refined by MD calculations, which included DPC micelles and explicit water. In the modelled complex the macrocycle of PMX-B is centered on the phosphate group at GlcN-B and additional contacts from polar side chains are formed to GlcN-A and Kdo-C, while hydrophobic side chains target the acyl region.

Our experimental data now confirm the model of the PMX-B LPS complex, proposed by Pristovsek et al[10]. Their investigations of the conformational behavior of PMX-B and PMX-B in water solution, free and bound to LPS, by homonuclear NMR and molecular modeling methods reveal that free peptides exist in equilibria of fast exchanging conformations with local preferences for a distorted type II β-turn from residues 5-8, and/or a γ-turn in residue 10. In the bound conformation they form an envelope-like bent cycle, which forms an amphiphilic separation of the two hydrophobic residues 6 and 7 from the charged Dab residues 4, 5, 8, and 9 which cluster on the lower side of the cycle. The structure does not display any intramolecular H-bonding, including the prominent β- and γ-turns. While our work confirmed the contacts made with moieties of the GlcN units proposed in the model complex from Pristovsek it is different in that it emphasizes the importance of contacts to Kdo units, in particular to Kdo-C.

In the complex of polymyxin M with LPS the bend in the heptacycle was in the opposite direction[126]. This resulted in a less dramatic separation of hydrophobic and hydrophilic side chains. In PMX-M position 7 is occupied by Thr, a much more polar residue. According to the data from our work, PMX-M may be slightly differently oriented, possibly forming some-what stronger contacts to the Kdo units. In the PMX-M—LPS complex electrostatic or polar interactions involving the Dab residues of polymyxin and the carbohydrate moieties of GlcN-B or Kdo-C and to a smaller extent, Kdo-D dominate, while in the case of PMX–B or –E additional hydrophobic interactions from sidechains of Leu-7 are likely to contribute to binding.

1.6 References

1. Quiocho, F.A., *Carbohydrate-binding proteins: tertiary structures and protein-sugar interactions.* Annu Rev Biochem, 1986. **55**: p. 287-315.
2. Karlsson, K.A., *Bacterium-host protein-carbohydrate interactions and pathogenicity.* Biochem Soc Trans, 1999. **27**(4): p. 471-4.
3. Lundquist, J.J. and E.J. Toone, *The cluster glycoside effect.* Chem Rev, 2002. **102**(2): p. 555-78.
4. Dunitz, J.D., *Win some, lose some: enthalpy-entropy compensation in weak intermolecular interactions.* Chem Biol, 1995. **2**(11): p. 709-12.
5. Weis, W.I. and K. Drickamer, *Structural basis of lectin-carbohydrate recognition.* Annu Rev Biochem, 1996. **65**: p. 441-73.
6. Lis, H. and N. Sharon, *Lectins: Carbohydrate-Specific Proteins That Mediate Cellular Recognition.* Chem Rev, 1998. **98**(2): p. 637-674.
7. Quiocho, F.A., *Probing the atomic interactions between proteins and carbohydrates.* Biochem Soc Trans, 1993. **21**(2): p. 442-8.
8. Bourne, Y., P. Rouge, and C. Cambillau, *X-ray structure of a biantennary octasaccharide-lectin complex refined at 2.3-A resolution.* J Biol Chem, 1992. **267**(1): p. 197-203.
9. Levitt, M. and M.F. Perutz, *Aromatic rings act as hydrogen bond acceptors.* J Mol Biol, 1988. **201**(4): p. 751-4.
10. Pristovsek, P. and J. Kidric, *Solution structure of polymyxins B and E and effect of binding to lipopolysaccharide: an NMR and molecular modeling study.* J Med Chem, 1999. **42**(22): p. 4604-13.
11. Schindler, M. and M.J. Osborn, *Interaction of divalent cations and polymyxin B with lipopolysaccharide.* Biochemistry, 1979. **18**(20): p. 4425-30.
12. Toone, E.I., *Structure and energetics of protein-carbohydrate complexes.* Current Opinion in Structural biology, 1994. **4**: p. 719 -728.
13. Bryce, R.A., I.H. Hillier, and J.H. Naismith, *Carbohydrate-protein recognition: molecular dynamics simulations and free energy analysis of oligosaccharide binding to concanavalin A.* Biophys J, 2001. **81**(3): p. 1373-88.
14. Carlomagno, T., *Ligand-target interactions: what can we learn from NMR?* Annu Rev Biophys Biomol Struct, 2005. **34**: p. 245-66.
15. Jim´enez-Barberob, A.P.a.J., *NMR studies of carbohydrate–protein interactions in solution.* Chemical Society Reviews, 1998. **27**: p. 133-143.

16. Otting, G., *Experimental NMR techniques for studies of protein-ligand interactions.* Current Opinion in Structural Biology, 1993. **3**: p. 760-768.
17. Wishart, D.S., B.D. Sykes, and F.M. Richards, *Relationship between nuclear magnetic resonance chemical shift and protein secondary structure.* J Mol Biol, 1991. **222**(2): p. 311-33.
18. Wishart, D.S. and B.D. Sykes, *The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data.* J Biomol NMR, 1994. **4**(2): p. 171-80.
19. Johnson, M.A. and B.M. Pinto, *NMR spectroscopic and molecular modeling studies of protein-carbohydrate and protein-peptide interactions.* Carbohydr Res, 2004. **339**(5): p. 907-28.
20. Neuhaus, D.a.W., M *The Nuclear Overhauser Effect,* . 1989: VCH Publishers, New York.
21. Wüthrich, K., *NMR of Proteins and Nucleic Acids*. 1986: Wiley, New York.
22. Lian , L.Y., , and Roberts,G.C.K., *practical approaches in NMR of biological macromolecules* 1993: Oxford university press 153-82.
23. Shuker, S.B., et al., *Discovering high-affinity ligands for proteins: SAR by NMR.* Science, 1996. **274**(5292): p. 1531-4.
24. Meyer, M.M.a.B., *Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy.* Angewandte chemie 1999. **38**(12): p. 1784-1788.
25. Meyer, B. and T. Peters, *NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors.* Angew Chem Int Ed Engl, 2003. **42**(8): p. 864-90.
26. Groves, P, Kover, K. E, Andre, S, Bandorowicz-Pikula, J, Batta, G, Bruix, M, Buchet, R, Canales, A, Canada, F. J, Gabius, H. J, Laurents, D. V, Naranjo, J. R, Palczewska, M, Pikula, S, Rial, E, Strzelecka-Kiliszek, A, Jimenez-Barbero, J and *Temperature dependence of ligand-protein complex formation as reflected by saturation transfer difference NMR experiments.* Magn Reson Chem 2007. **9**: p. 745–748.
27. Peters, M.V.a.T., *Application of NMR Based Binding Assays to Identify Key Hydroxy Groups for Intermolecular Recognition.* J. Am. Chem. Soc, 2000. **122**: p. 6093-6099.
28. Dalvit, C., et al., *Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water.* J Biomol NMR, 2000. **18**(1): p. 65-8.
29. Angulo, J., et al., *NMR analysis of carbohydrate-protein interactions.* Methods Enzymol, 2006. **416**: p. 12-30.
30. Vincent, S.J., C. Zwahlen, and G. Bodenhausen, *Suppression of spin diffusion in selected frequency bands of nuclear Overhauser spectra.* J Biomol NMR, 1996. **7**(2): p. 169-72.
31. Vincent, S.J., et al., *The conformation of NAD+ bound to lactate dehydrogenase determined by nuclear magnetic resonance with suppression of spin diffusion.* Proc Natl Acad Sci U S A, 1997. **94**(9): p. 4383-8.
32. Murali, N., et al., *Two-dimensional transferred nuclear Overhauser effect spectroscopy (TRNOESY) studies of nucleotide conformations in creatine kinase complexes: effects due to weak nonspecific binding.* Biochemistry, 1993. **32**(47): p. 12941-8.
33. Weimar, T., et al., *Transferred nuclear Overhauser enhancement experiments show that the monoclonal antibody strep 9 selects a local minimum*

conformation of a Streptococcus group A trisaccharide-hapten. Biochemistry, 1995. **34**(41): p. 13672-81.
34. Arepalli, S.R., et al., *Identification of protein-mediated indirect NOE effects in a disaccharide-Fab' complex by transferred ROESY.* J Magn Reson B, 1995. **106**(2): p. 195-8.
35. Kar, L., et al., *Specificity and affinity of binding of phosphate-containing compounds to CheY protein.* Biochem J, 1992. **287 (Pt 2)**: p. 533-43.
36. Odaka, A., et al., *Isotope-edited nuclear magnetic resonance study of Fv fragment of anti-dansyl mouse monoclonal antibody: recognition of the dansyl hapten.* Biochemistry, 1992. **31**(44): p. 10686-91.
37. Miller, A.F., M.Z. Papastavros, and A.G. Redfield, *NMR studies of the conformational change in human N-p21ras produced by replacement of bound GDP with the GTP analog GTP gamma S.* Biochemistry, 1992. **31**(42): p. 10208-16.
38. Tolman, J.R., et al., *Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution.* Proc Natl Acad Sci U S A, 1995. **92**(20): p. 9279-83.
39. Bursulaya, B.D., et al., *Comparative study of several algorithms for flexible ligand docking.* J Comput Aided Mol Des, 2003. **17**(11): p. 755-63.
40. Imberty, A., et al., *Molecular modelling of protein-carbohydrate interactions. Docking of monosaccharides in the binding site of concanavalin A.* Glycobiology, 1991. **1**(6): p. 631-42.
41. Taroni, C., S. Jones, and J.M. Thornton, *Analysis and prediction of carbohydrate binding sites.* Protein Eng, 2000. **13**(2): p. 89-98.
42. Stoddard, B.L. and G.K. Farber, *Direct measurement of reactivity in the protein crystal by steady-state kinetic studies.* Structure, 1995. **3**(10): p. 991-6.
43. Schlichting, I., *X-ray crystallography of protein-ligand interactions.* Methods Mol Biol, 2005. **305**: p. 155-66.
44. Jameson, D.M., J.C. Croney, and P.D. Moens, *Fluorescence: basic concepts, practical aspects, and some anecdotes.* Methods Enzymol, 2003. **360**: p. 1-43.
45. Clegg, R.M., *Fluorescence resonance energy transfer.* Curr Opin Biotechnol, 1995. **6**(1): p. 103-10.
46. Dam, T.K. and C.F. Brewer, *Multivalent protein-carbohydrate interactions: isothermal titration microcalorimetry studies.* Methods Enzymol, 2004. **379**: p. 107-28.
47. Rich, R.L. and D.G. Myszka, *Survey of the year 2005 commercial optical biosensor literature.* J Mol Recognit, 2006. **19**(6): p. 478-534.
48. Wyer, J.R., et al., *T cell receptor and coreceptor CD8 alphaalpha bind peptide-MHC independently and with distinct kinetics.* Immunity, 1999. **10**(2): p. 219-25.
49. Willcox, B.E., et al., *TCR binding to peptide-MHC stabilizes a flexible recognition interface.* Immunity, 1999. **10**(3): p. 357-65.
50. Piliarik, M., H. Vaisocherova, and J. Homola, *Surface plasmon resonance biosensing.* Methods Mol Biol, 2009. **503**: p. 65-88.
51. Smith, E.A., et al., *Surface plasmon resonance imaging studies of protein-carbohydrate interactions.* J Am Chem Soc, 2003. **125**(20): p. 6140-8.
52. Pfeiffer, R., *Untersuchungen über das Choleragift.* Medical Microbiology and Immunology. **11**: p. 393–412.
53. Rietschel, E.T. and H. Brade, *Bacterial endotoxins.* Sci Am, 1992. **267**(2): p. 54-61.

54. Vogel, S.N.a.H., M. M., *Role of cytokines in endotoxin-mediated host responses.(J. J.Oppenheim and E. M. Shevach, eds)*. 1990: Oxford University Press, New York. 238–258.
55. Heine, H., E.T. Rietschel, and A.J. Ulmer, *The biology of endotoxin.* Mol Biotechnol, 2001. **19**(3): p. 279-96.
56. Raetz, C.R., *Biochemistry of endotoxins.* Annu Rev Biochem, 1990. **59**: p. 129-70.
57. Qureshi, N., et al., *Position of ester groups in the lipid A backbone of lipopolysaccharides obtained from Salmonella typhimurium.* J Biol Chem, 1983. **258**(21): p. 12947-51.
58. Chang, C.C. and Nowotny, *Relation of structure to function in bacterial O-antigens--VII. Endotoxicity of 'lipid A'.* Immunochemistry, 1975. **12**(1): p. 19-28.
59. van Alphen, L., et al., *31P nuclear magnetic resonance and freeze-fracture electron microscopy studies on Escherichia coli. II. Lipopolysaccharide and lipopolysaccharide-phospholipid complexes.* Biochim Biophys Acta, 1980. **597**(3): p. 502-17.
60. Strain, S.M., S.W. Fesik, and I.M. Armitage, *Structure and metal-binding properties of lipopolysaccharides from heptoseless mutants of Escherichia coli studied by 13C and 31P nuclear magnetic resonance.* J Biol Chem, 1983. **258**(22): p. 13466-77.
61. Bush, C.A., M. Martin-Pastor, and A. Imberty, *Structure and conformation of complex carbohydrates of glycoproteins, glycolipids, and bacterial polysaccharides.* Annu Rev Biophys Biomol Struct, 1999. **28**: p. 269-93.
62. Rana, F.R., et al., *Interactions between magainin 2 and Salmonella typhimurium outer membranes: effect of lipopolysaccharide structure.* Biochemistry, 1991. **30**(24): p. 5858-66.
63. Ferguson, A.D., et al., *Siderophore-mediated iron transport: crystal structure of FhuA with bound lipopolysaccharide.* Science, 1998. **282**(5397): p. 2215-20.
64. Wang, W., et al., *Structure and dynamics of 13C,15N-labeled lipopolysaccharides in a membrane mimetic.* Angew Chem Int Ed Engl, 2008. **47**(51): p. 9870-4.
65. Garcia Vescovi, E., F.C. Soncini, and E.A. Groisman, *Mg2+ as an extracellular signal: environmental regulation of Salmonella virulence.* Cell, 1996. **84**(1): p. 165-74.
66. Loppnow, H., et al., *Induction of human interleukin 1 by bacterial and synthetic lipid A.* Eur J Immunol, 1986. **16**(10): p. 1263-7.
67. Kovach, N.L., et al., *Lipid IVA inhibits synthesis and release of tumor necrosis factor induced by lipopolysaccharide in human whole blood ex vivo.* J Exp Med, 1990. **172**(1): p. 77-84.
68. Galanos, C., et al., *Biological activity of synthetic heptaacyl lipid A representing a component of Salmonella minnesota R595 lipid A.* Eur J Biochem, 1986. **160**(1): p. 55-9.
69. Christ, W.J., et al., *E5531, a pure endotoxin antagonist of high potency.* Science, 1995. **268**(5207): p. 80-3.
70. Seydel, U., et al., *Phase behavior, supramolecular structure, and molecular conformation of lipopolysaccharide.* Immunobiology, 1993. **187**(3-5): p. 191-211.
71. Takayama, K., et al., *Monomeric Re lipopolysaccharide from Escherichia coli is more active than the aggregated form in the Limulus amebocyte lysate assay*

and in inducing Egr-1 mRNA in murine peritoneal macrophages. J Biol Chem, 1994. **269**(3): p. 2241-4.
72. Ingalls, R.R. and D.T. Golenbock, *CD11c/CD18, a transmembrane signaling receptor for lipopolysaccharide.* J Exp Med, 1995. **181**(4): p. 1473-9.
73. Dobrovolskaia, M.A. and S.N. Vogel, *Toll receptors, CD14, and macrophage activation and deactivation by LPS.* Microbes Infect, 2002. **4**(9): p. 903-14.
74. Triantafilou, M. and K. Triantafilou, *Lipopolysaccharide recognition: CD14, TLRs and the LPS-activation cluster.* Trends Immunol, 2002. **23**(6): p. 301-4.
75. Harder, T. and K. Simons, *Caveolae, DIGs, and the dynamics of sphingolipid-cholesterol microdomains.* Curr Opin Cell Biol, 1997. **9**(4): p. 534-42.
76. Brown, D.A. and J.K. Rose, *Sorting of GPI-anchored proteins to glycolipid-enriched membrane subdomains during transport to the apical cell surface.* Cell, 1992. **68**(3): p. 533-44.
77. Cinek, T. and V. Horejsi, *The nature of large noncovalent complexes containing glycosyl-phosphatidylinositol-anchored membrane glycoproteins and protein tyrosine kinases.* J Immunol, 1992. **149**(7): p. 2262-70.
78. Gutsmann, T., et al., *Dual role of lipopolysaccharide (LPS)-binding protein in neutralization of LPS and enhancement of LPS-induced activation of mononuclear cells.* Infect Immun, 2001. **69**(11): p. 6942-50.
79. Weiss, J., et al., *Purification and characterization of a potent bactericidal and membrane active protein from the granules of human polymorphonuclear leukocytes.* J Biol Chem, 1978. **253**(8): p. 2664-72.
80. Gazzano-Santoro, H., et al., *Competition between rBPI23, a recombinant fragment of bactericidal/permeability-increasing protein, and lipopolysaccharide (LPS)-binding protein for binding to LPS and gram-negative bacteria.* Infect Immun, 1994. **62**(4): p. 1185-91.
81. Juan, T.S., et al., *Soluble CD14 truncated at amino acid 152 binds lipopolysaccharide (LPS) and enables cellular response to LPS.* J Biol Chem, 1995. **270**(3): p. 1382-7.
82. Stelter, F., et al., *Differential impact of substitution of amino acids 9-13 and 91-101 of human CD14 on soluble CD14-dependent activation of cells by lipopolysaccharide.* J Immunol, 1999. **163**(11): p. 6035-44.
83. Voss, S., et al., *A CD14 domain with lipopolysaccharide-binding and -neutralizing activity.* Chembiochem, 2006. **7**(2): p. 275-86.
84. Pugin, J., et al., *CD14 is a pattern recognition receptor.* Immunity, 1994. **1**(6): p. 509-16.
85. Heumann, D., M.P. Glauser, and T. Calandra, *Molecular basis of host-pathogen interaction in septic shock.* Curr Opin Microbiol, 1998. **1**(1): p. 49-55.
86. Kol, A., et al., *Cutting edge: heat shock protein (HSP) 60 activates the innate immune response: CD14 is an essential receptor for HSP60 activation of mononuclear cells.* J Immunol, 2000. **164**(1): p. 13-7.
87. Schmitz, G. and E. Orso, *CD14 signalling in lipid rafts: new ligands and co-receptors.* Curr Opin Lipidol, 2002. **13**(5): p. 513-21.
88. Wurfel, M.M., E. Hailman, and S.D. Wright, *Soluble CD14 acts as a shuttle in the neutralization of lipopolysaccharide (LPS) by LPS-binding protein and reconstituted high density lipoprotein.* J Exp Med, 1995. **181**(5): p. 1743-54.
89. Yu, B., E. Hailman, and S.D. Wright, *Lipopolysaccharide binding protein and soluble CD14 catalyze exchange of phospholipids.* J Clin Invest, 1997. **99**(2): p. 315-24.

90. Troelstra, A., et al., *Dual effects of soluble CD14 on LPS priming of neutrophils.* J Leukoc Biol, 1997. **61**(2): p. 173-8.
91. Kim, J.I., et al., *Crystal structure of CD14 and its implications for lipopolysaccharide signaling.* J Biol Chem, 2005. **280**(12): p. 11347-51.
92. Wu, Z., et al., *Chicken CD14, unlike mammalian CD14, is trans-membrane rather than GPI-anchored.* Dev Comp Immunol, 2009. **33**(1): p. 97-104.
93. Lemaitre, B., et al., *The dorsoventral regulatory gene cassette spatzle/Toll/cactus controls the potent antifungal response in Drosophila adults.* Cell, 1996. **86**(6): p. 973-83.
94. Medzhitov, G.M.B.a.R., *Toll-Like Receptor Signaling Pathways.* Science, 2003. **300**: p. 1524.
95. Rietschel, E.T., et al., *Bacterial endotoxin: molecular relationships of structure to activity and function.* Faseb J, 1994. **8**(2): p. 217-25.
96. Munford, R.S. and C.L. Hall, *Purification of acyloxyacyl hydrolase, a leukocyte enzyme that removes secondary acyl chains from bacterial lipopolysaccharides.* J Biol Chem, 1989. **264**(26): p. 15613-9.
97. Andersson, J., Melchers, F., Galanos, C., and Luderitz. O., *The mitogenic effect of lipopolysaccharide on bone marrow-derived mouse lymphocytes.Lipid A as the mitogenic part of the molecule.* J. Exp. Med, 1973. **137 .** p. 943–953.
98. Mattern, T., et al., *Stimulation of human T lymphocytes by LPS is MHC unrestricted, but strongly dependent on B7 interactions.* J Immunol, 1998. **160**(7): p. 3412-8.
99. Vogel, S.N., M.L. Hilfiker, and M.J. Caulfield, *Endotoxin-induced T lymphocyte proliferation.* J Immunol, 1983. **130**(4): p. 1774-9.
100. Pober, J.S. and R.S. Cotran, *The role of endothelial cells in inflammation.* Transplantation, 1990. **50**(4): p. 537-44.
101. Lazaron, V. and D.L. Dunn, *Molecular biology of endotoxin antagonism.* World J Surg, 2002. **26**(7): p. 790-8.
102. Appelmelk BJ, C.J., *The protective role of antibodies to the lipopolysaccharide core region.* Vol. Vol. II. 1992: Boca Raton: CRC Press. 375–410.
103. Warren, H.S., et al., *Binding and neutralization of endotoxin by Limulus antilipopolysaccharide factor.* Infect Immun, 1992. **60**(6): p. 2506-13.
104. Maliszewski, C.R., *CD14 and immune response to lipopolysaccharide.* Science, 1991. **252**(5010): p. 1321-2.
105. Jerala, R. and M. Porro, *Endotoxin neutralizing peptides.* Curr Top Med Chem, 2004. **4**(11): p. 1173-84.
106. Bulet, P., R. Stocklin, and L. Menin, *Anti-microbial peptides: from invertebrates to vertebrates.* Immunol Rev, 2004. **198**: p. 169-84.
107. Mygind, P.H., et al., *Plectasin is a peptide antibiotic with therapeutic potential from a saprophytic fungus.* Nature, 2005. **437**(7061): p. 975-80.
108. Shai, Y., *Mechanism of the binding, insertion and destabilization of phospholipid bilayer membranes by alpha-helical antimicrobial and cell non-selective membrane-lytic peptides.* Biochim Biophys Acta, 1999. **1462**(1-2): p. 55-70.
109. Park, C.B., H.S. Kim, and S.C. Kim, *Mechanism of action of the antimicrobial peptide buforin II: buforin II kills microorganisms by penetrating the cell membrane and inhibiting cellular functions.* Biochem Biophys Res Commun, 1998. **244**(1): p. 253-7.

110. Hong, S.Y., J.E. Oh, and K.H. Lee, *Effect of D-amino acid substitution on the stability, the secondary structure, and the activity of membrane-active peptide.* Biochem Pharmacol, 1999. **58**(11): p. 1775-80.
111. Yosef Rosenfeld and Y. Shai, *Lipopolysaccharide (Endotoxin)-host defense antibacterial peptides interactions: Role in bacterial resistance and prevention of sepsis.* Biochimica et Biophysica Acta, 2006.
112. Stansly, P.G. and M.E. Schlosser, *Studies on Polymyxin: Isolation and Identification of Bacillus polymyxa and Differentiation of Polymyxin from Certain Known Antibiotics.* J Bacteriol, 1947. **54**(5): p. 549-56.
113. Koike, M., K. Iida, and T. Matsuo, *Electron microscopic studies on mode of action of polymyxin.* J Bacteriol, 1969. **97**(1): p. 448-52.
114. Lopes, J. and W.E. Inniss, *Electron microscopy of effect of polymyxin on Escherichia coli lipopolysaccharide.* J Bacteriol, 1969. **100**(2): p. 1128-9.
115. Lerner, R.G., et al., *Stimulation of human leukocyte thromboplastic activity by endotoxin.* Proc Soc Exp Biol Med, 1971. **138**(1): p. 145-8.
116. Bannatyne, R.M., et al., *Inhibition of the biologic effects of endotoxin on neutrophils by polymyxin B sulfate.* J Infect Dis, 1977. **136**(4): p. 469-74.
117. Corrigan, J.J., Jr. and B.M. Bell, *Endotoxin-induced intravascular coagulation: prevention with polymyxin B sulfate.* J Lab Clin Med, 1971. **77**(5): p. 802-10.
118. Appelmelk, B.J., et al., *Polymyxin B-horseradish peroxidase conjugates as tools in endotoxin research.* Anal Biochem, 1992. **207**(2): p. 311-6.
119. Morrison, D.C. and D.M. Jacobs, *Binding of polymyxin B to the lipid A portion of bacterial lipopolysaccharides.* Immunochemistry, 1976. **13**(10): p. 813-8.
120. Koch-Weser, J., et al., *Adverse effects of sodium colistimethate. Manifestations and specific reaction rates during 317 courses of therapy.* Ann Intern Med, 1970. **72**(6): p. 857-68.
121. Falagas, M.E. and S.K. Kasiakou, *Toxicity of polymyxins: a systematic review of the evidence from old and recent studies.* Crit Care, 2006. **10**(1): p. R27.
122. Storm, D.R., K.S. Rosenthal, and P.E. Swanson, *Polymyxin and related peptide antibiotics.* Annu Rev Biochem, 1977. **46**: p. 723-63.
123. David, S.A., *Towards a rational development of anti-endotoxin agents: novel approaches to sequestration of bacterial endotoxins with small molecules.* J Mol Recognit, 2001. **14**(6): p. 370-87.
124. Cohen, J. and J.S. McConnell, *Antibiotic-induced endotoxin release.* Lancet, 1985. **2**(8463): p. 1069-70.
125. Prins, J.M., et al., *Release of tumor necrosis factor alpha and interleukin 6 during antibiotic killing of Escherichia coli in whole blood: influence of antibiotic class, antibiotic concentration, and presence of septic serum.* Infect Immun, 1995. **63**(6): p. 2236-42.
126. Martin, N.I., et al., *Isolation, structural characterization, and properties of mattacin (polymyxin M), a cyclic peptide antibiotic produced by Paenibacillus kobensis M.* J Biol Chem, 2003. **278**(15): p. 13124-32.

Chapter 2

2.0 Introduction

G protein-coupled receptors, also known as GPCRs, represent a very large sub family of receptors[1], which detect a wide range of signaling molecules outside the cell and, in response, turn on signaling cascades inside the cell. These integral membrane proteins control countless processes in the body, including appetite, vision, smell, and heartbeat, endowing this receptor class with enormous therapeutic potential. Nearly 800 such receptors are thought to be present in the human genome[2]. Many diseases are linked to GPCRs. Not surprisingly half of the drug targets in the pharmaceutical industry are GPCRs[3]. Since the cloning of the first receptors, extensive experimental work has uncovered multiple aspects of their function and challenged many traditional paradigms.

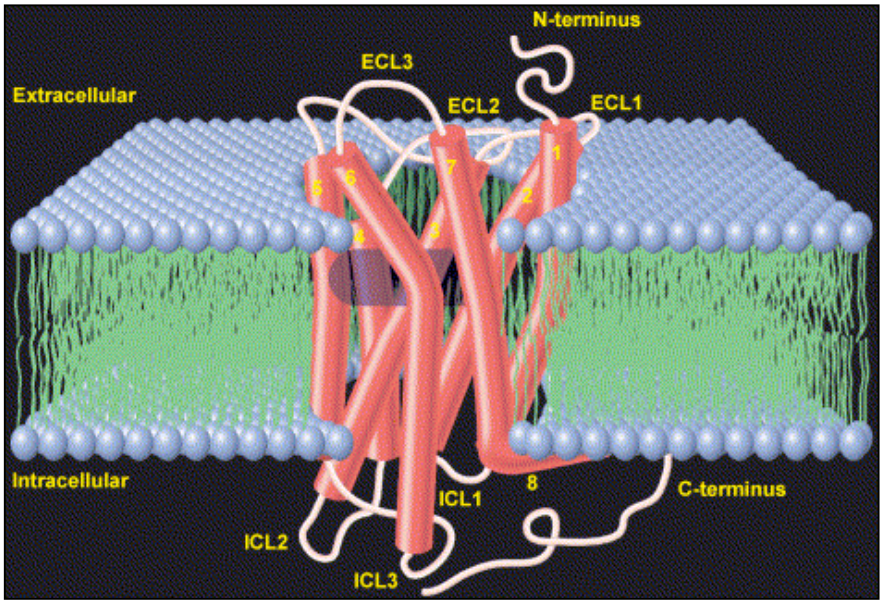


Figure 1: Schematic representation of GPCR[4].

Despite the remarkable structural diversity of natural GPCR agonists and low sequence homology among GPCRs, hydropathy analysis and biochemical data suggest that all GPCRs share a common molecular architecture. (Figure1). This raises the question, what structural features enable GPCRs to recognize and bind diffusible ligands? How structurally conserved are they within the same class and what is the

importance of their similarities and differences from a functional point of view .Thus a critical need exists for structural information to understand the mechanism of receptor activation and regulation and enormous effort has been expended to obtain structural details for GPCRs.

Despite the extensive efforts, only in the past year great strides in the field of GPCR structural biology have been achieved with high resolution crystal structures of three new members being published[5-8]. Structural studies of intact GPCRs remain a challenging problem because of the daunting technical difficulties in expression, purification and the problems associated with their structural analysis. When using NMR the large size of the GPCR /micellar complex, poor chemical shift dispersion and the redundancy of the residues found in the transmembrane region of the receptor complicate the analysis. An alternative strategy of production of smaller fragments of GPCRs helps to overcome the problems associated with protein expression, and also aids in their spectroscopic characterization due to simplification of spectra.

2.1 A brief history of GPCRs

The grandparents of modern receptor biology are probably Ehrlich and Langley. While Ehrlich was not exactly studying what we consider to be receptors in the modern sense, his “side chain theory” attempted to explain how antigens bound to cells. His conclusion was “corpora non agunt nisi fixata” – agents cannot act unless they are bound – a statement frequently quoted in the context of modern receptor biology[9]. However, contained in Langley’s study of the neuromuscular junction, was the first reference to a “receptive substance” describing the cellular sites of interaction of drugs curare/nicotine and atropine/pilocarpine[10]. The discovery of cyclic AMP and adenylyl cyclase by Sutherland marks the historical beginning of study of G-protein coupled receptors[11, 12]. In 1971, Rodbell proposed the existence of a guanine nucleotide regulatory protein that serves as a transducer between hormone receptors and adenylyl cyclase[13]. Gilman and colleagues subsequently demonstrated the existence of this protein[14] and later purified the protein and named it Gs[15]. Thus hydrolysis of GTP was found to allow heterotrimeric G-proteins to couple receptors and to activate or inhibit enzymes and ion channels, allowing modulation of cellular physiology by external agents[15-19]. Taken together, these discoveries provided the

first chain of molecular events that convey signaling from the outside to the inside of the cell. The era of molecular research on the receptors (studies that assess receptor properties directly, rather than inferring them from downstream effector function) can reasonably be dated to 1970. Until recently, our atomic-level understanding of GPCRs has been based on rhodopsin (2000) in its inactive state[20] After 7 years, the drought has been ended with a flurry of papers around the structures of beta adrenergic receptors, the target of beta blockers[5-7, 21, 22], human A_{2A} adenosine receptor[7, 8] and significant inroads made in resolving the activated state of bovine rhodopsin[23, 24]. Together these efforts represent the first examples of what is sure to be a blossoming of information for this important class of membrane proteins.

2.2 GPCR signaling

Much of vertebrate physiology is based on GPCR signal transduction. As receptors for hormones, neurotransmitters, ions, photons and other stimuli, GPCRs are among the essential nodes of communication between the internal and external environments of cells. The classical role of GPCRs is to couple the binding of agonists to the activation of specific heterotrimeric G proteins, leading to the modulation of downstream effector proteins. Natural and synthetic ligands can be grouped into different efficacy classes. Full agonists are capable of maximal receptor stimulation; partial agonists are unable to elicit full activity even at saturating concentrations; neutral antagonists have no effect on signaling activity but can prevent other ligands from binding to the receptor; and inverse agonists reduce the level of basal or constitutive activity below that of the unliganded receptor.

A wide variety of ligands, including biogenic amines, amino acids, ions, lipids, peptides and proteins, use GPCRs to stimulate various cytoplasmic and nuclear targets and these ligands vary in size from small molecules to peptides to large proteins. A general model of GPCR activation postulates that GPCRs are in equilibrium between an active state and an inactive state which differs by disposition of transmembrane helices, turns, cytoplasmic domains and the agonists act by stabilizing the activated state[25].

GPCRs owe their name to their extensively studied interaction with heterotrimeric G proteins (composed of an α-, β- and γ-subunit). When a ligand binds to the GPCR it causes a conformational change in the GPCR which allows it to act as a guanine nucleotide exchange factor (GEF). The GPCR can then activate an associated G-protein by exchanging its bound GDP for a GTP. The G-protein's α subunit, together with the bound GTP, can then dissociate from the β and γ subunits to further affect intracellular signaling proteins or target functional proteins directly depending on the α subunit type (G_s ,G_i, G_q, G_{12}). Consequently, the Gα- and Gβγ- subunits stimulate effector molecules, which include adenylyl and guanylyl cyclases, phosphodiesterases,phospholipaseA2 (PLA2), phospholipaseC (PLC) and phospho inositide3-kinases (PI3Ks), thereby activating or inhibiting the production of a variety of second messengers such as cAMP, cGMP, diacylglycerol, inositol(1,4,5)-trisphosphate (IP3), phosphatidyl inositol(3,4,5)-trisphosphate [PIP3], arachidonic acid and phosphatidic acid, in addition to promoting increases in the intracellular concentration of Ca^{2+} and the opening or closing of a variety of ion channels (Figure 2). It results in the rapid effects on hormone secretion, muscle contraction and a variety of other physiological functions. Long terms changes in gene expression are mediated by phosphorylation of transcription factors.

However, beyond the dogma of 'GPCR–G-protein', in the late 1990s, evidence began accumulating to suggest that some GPCRs are able to signal without G-proteins. For example, the adaptor molecule arrestin binds many phosphorylated GPCRs and is primarily involved in targeting these receptors for endocytosis. In kidney cells, the bradykinin receptor B2 has been shown to interact directly with a protein tyrosine phosphatase. Other structural features, which include PDZ, SH2 and SH3 domains, and polyproline-containing regions also provide the molecular basis for direct interactions between GPCRs and several intracellular signaling molecules[26]. Furthermore, it has been shown that a variety of cellular responses are mediated by the activation of effector molecules through novel molecular mechanisms, many of which do not involve the stimulation of classical second messengers.

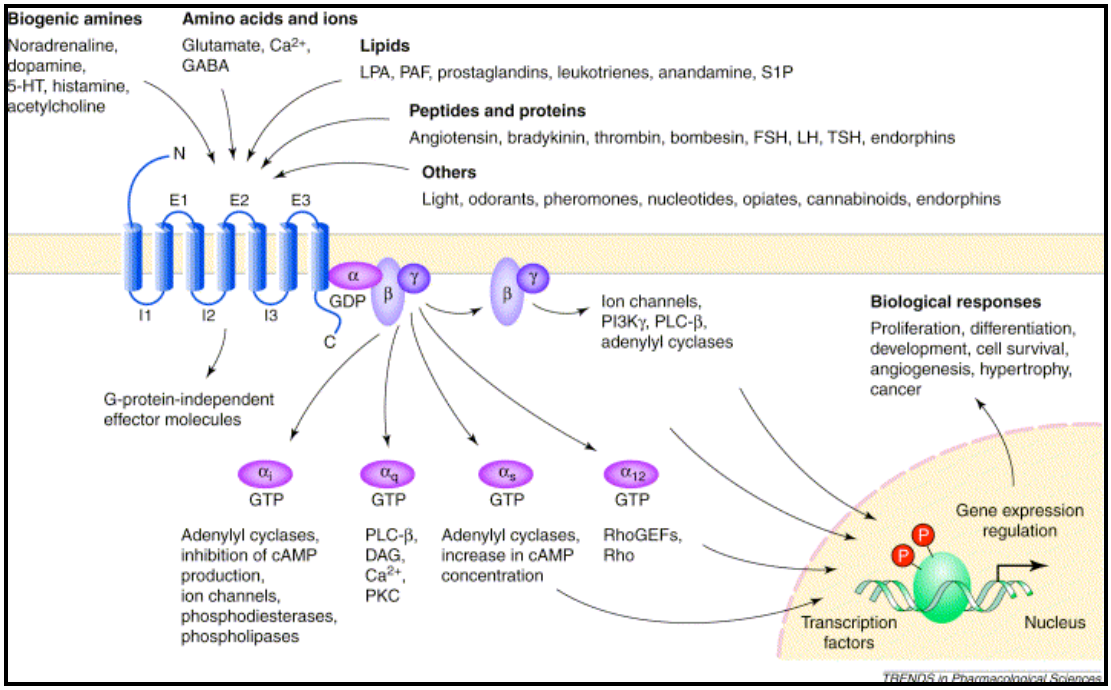


Figure 2: Diversity of G-protein-coupled receptors (GPCRs)[27].

Abbreviations: DAG, diacylglycerol; FSH, follicle-stimulating hormone; GEF, guanine nucleotide exchange factor; LH, leuteinizing hormone; LPA, lysophosphatidic acid; PAF, platelet-activating factor; PI3K, phosphoinositide 3-kinase; PKC, protein kinase C; PLC, phospholipase C; S1P, sphingosine-1-phosphate; TSH, thyroid-stimulating hormone.

GPCRs become desensitized when exposed to their ligand for a prolonged period of time. There are two recognized forms of desensitization:1) homologous desensitization, in which the activated GPCR is downregulated; and 2) heterologous desensitization, wherein the activated GPCR causes downregulation of a different GPCR. The key reaction of this downregulation is the phosphorylation of the intracellular (or cytoplasmic) receptor domain by protein kinases. Thus the final outcome of most biological responses mediated by GPCRs are not dependent on a single biochemical route, but results from the integration of a complex network of biochemical responses.

2.3 Significance of GPCRs

How do we communicate with the outside world? How are our senses of vision, smell, taste and pain controlled at the cellular and molecular levels? What causes medical conditions like allergies, hypertension, depression, obesity and various central nervous system disorders? G-protein coupled receptors (GPCRs) provide a major part of the answer to all of these questions. They are involved in the control of several

important aspects of our behavior and physiology. Of the 720–800 human GPCRs, which accounts for about 2% of the human genome, 380 were found to be unique functional non-olfactory/non sensory GPCR sequences, for which endogenous ligands are expected. They are referred to as endo-GPCRs[28, 29]. These receptors are expressed in different tissues and regulate various aspects of our physiology. The endo-GPCR group has attracted much attention in recent years. Examples are receptors for hormones such as calcitonin and luteinizing hormone or neurotransmitters such as serotonin and dopamine.

There is a growing body of evidence that many mutated forms of GPCRs are associated with a wide spectrum of disease phenotypes and predispositions. Mutations that cause inactive receptor proteins are often referred to as loss-of-function (LOF) mutations. By contrast, constitutively active mutants (CAMs) result in autonomous signaling in the absence of agonist. These mutations results in various pathological processes which lead to numerous diseases, including cardiovascular, mental disorders, retinal degeneration, cancer, AIDS etc. The investigation of these mutations gives insight into the causes of human genetic disease and provides perspective on strategies for drug discovery that take into account the potential for the development of drugs targeted at mutated and wild-type GPCRs[30, 31], [32].

2.4 Classification of GPCRs

The GPCR superfamily is historically classified into three main families[33]. Rhodopsin belongs to family 1 and is the most intensively studied GPCR, so family 1 is also called the rhodopsin-like family. The other two main subfamilies are the secretin-like receptor (family 2) and the metabotropic glutamate-like receptor (family 3) as shown in Figure 3. Family 1 comprises almost 90% of all GPCRs and is by far the best-studied subfamily. The members of family 1 GPCRs, possess a characteristic disulfide bridge between extracellular loops E-I and E-II. Most of these receptors also have a palmitoylated cysteine in the carboxy-terminal tail, which serves as a membrane anchor. The crystal structure of rhodopsin has indicated that some of the transmembrane domains of family 1 receptors are ‘tilted’ and ‘kinked’ due to the presence of certain amino acids such as proline that distort the transmembrane helices. A short amphiphilic helix, lying nearly perpendicular to TM VII and therefore parallel

to the cytoplasmic membrane, is part of the C-terminal tail.

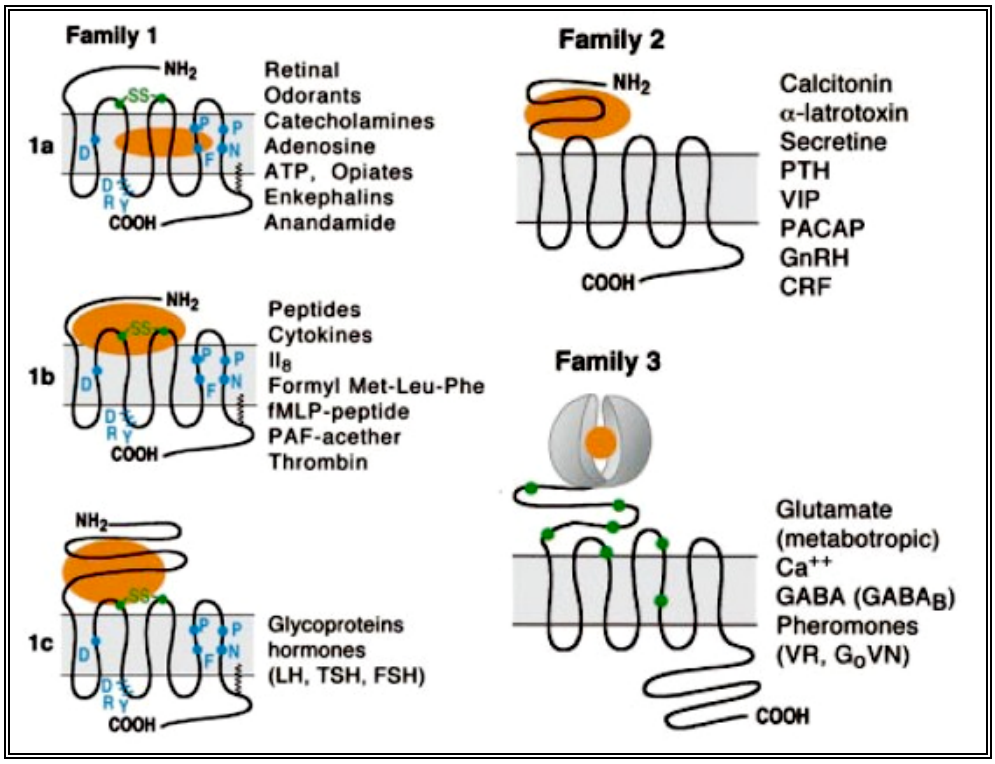


Figure 3: Three family classes of GPCRs.
Family 1 comprises the rhodopsin-like receptors, which are subdivided into three subfamilies.
Family 2 includes hormone receptors
Family 3 comprises the metabotropic glutamate receptors.
Sites known to be involved in ligand binding are highlighted in orange.

Family 1 can be subdivided into 3 subfamilies family 1a, 1b and 1c. Group la contains GPCRs such as rhodopsin, opioid receptors and ß-adrenergic receptors, which binds to small ligands. The binding site (orange coloured) is within the seven TM helices in the hydrophobic membrane compartment. Receptors of important neurotransmitters such as dopamine, serotonin, peptide hormones as well as neuropeptides, and in particular NPY are mainly members of family lb with the ligand binding site involving the N-terminal part and the extracellular loops. Group lc contains receptors for glycoprotein hormones and a large extracellular domain characterizes them. Family 2 GPCRs have been found in many animal species but not in plants, fungi or prokaryotes. Many members of this family contain two additional structural features in addition to the classical 7 TM region: i) a mucin-like region rich in serine and threonine residues; ii) a conserved cysteine-rich proteolysis domain[34]. They are morphologically very similar to family lc GPCRs although the sequence homology is very low and they lack the characteristic disulfide bond and palmitoylation site of

family 1 GPCRs. The ligands for the family 2 GPCRs are large peptides including the closely related 29-residue peptide glucagon, the 27-residue peptide secretin and the 28-residue vasoactive intestinal peptide (VIP), or the calcium and phosphate ion concentration regulating 32-residue peptide calcitonin and the 84 amino acid peptide parathyroid hormone (PTH).

The family 3 GPCRs are receptors for the main neurotransmitters glutamate and γ-aminobutyric acid, for Ca^{2+}, for sweet and amino acid taste compounds, and for some pheromone molecules, as well as for odorants in fish. Although none of these family 3 receptors have been found in plants, members have been identified in ancient organisms, such as slime molds (*Dictyostelium*) and sponges[35]. A long amino terminus and a long carboxy tail characterize these receptors. These receptors possess a very large extracellular domain constituted of two lobes (ligand binding regions LBR 1 and 2) that close like a Venus flytrap upon ligand binding. In spite of the low sequence similarity of family 3 receptors with that of rhodopsin-like receptors they share some structural similarities like 1) the conserved disulfide bond between E-I and E-II is also conserved in family 3 members, 2) it has an 8th helix (H8) after TM7 like in the rhodopsin structure. A unique characteristic of these receptors is that the third intracellular loop is short and highly conserved.

Fredriksson and colleagues divided 802 (known and predicted) human GPCRs into families on the basis of phylogenetic criteria. This system classifies most of the human GPCRs into five main families (Table.1) termed *Glutamate*, *Rhodopsin*, *Adhesion*, *Frizzled/Taste2* and *Secretin* (shortened to the acronym GRAFS)[28] The main difference between this nomenclature system and the former classification systems is the further division of family 2 into the *Secretin* family and the *Adhesion* family. Each family covers several subfamilies of GPCRs. In addition, some putative families, orphan receptors and non-GPCRs families are also included (refer to www.gpcr.org).

Property	Rhodopsin	Secretin	Adhesion	Glutamate	Frizzled	Taste2
Number of full-length receptor proteins*	672 (388 ORs)	15	33	22	11	25
Number of identified major drug targets‡	>39	4	0	3	0	0
Number of orphans*	63 (not including ORs)	0	30	7	0	21
Type of ligand	Peptides, proteins (including enzymes), small organic compounds, lipid-like substances, nucleotides	Peptides, proteins	Proteins, glycosaminoglycan	Amino acids, cations, small organic compounds, carbohydrates	Proteins	Small organic compounds
Extended N termini§	No	No	Yes	Yes	Yes	Yes
Conserved functional domains in the N termini§	No	No	Yes	Yes	Yes	Yes
Type of functional domains in the N termini (not including transmembrane)	Proteolytic sites, LRR, LDLa	HBD	GPS, TSP1, HBD, PTX, EGF, OLF, GBL, CA, LamG, EGF-Lam, LRR, Ig, SEA, Calx-beta, EAR, CUB, C-type lectin	VFTM, SUSHI	Wnt binding domain	–
Proteolytic processing of the N termini in family members	Yes	No	Yes	No	No	No
Conserved cysteine residues in EL1 and EL2‖	Yes	Yes	Yes	Yes	Yes	No
Suitable as drug targets?	Yes¶	Yes	Yes	Yes#	Yes	No

Table 1: A summary of properties for the G protein-coupled receptor (GPCR) families[36]

2.5 Expression of GPCRs

Recombinant overexpression of GPCRs is essential for their structural characterization and for the structure based drug design. Heterologous expression of functional GPCRs has been accomplished in *Escherichia coli*, yeast, insect and mammalian cells, but with varying degree of success because of the differences in host cell environment[37, 38] (Table 2). *E.coli* system definitely has the advantage in that it is simple, rapid to use, various expression plasmids and *E.coli* strains are available, scale-up costs are reasonable and uniform labeling of the protein for NMR is comparatively cheaper than with the other expression systems. The toxic effects of GPCRs on bacterial membrane, codon usage efficiency, stability of the expressed protein and the lack of post translational modification are the major setbacks of this system. Several methods have been developed to overcome these problems and increase the protein yield. Notably, fusing target proteins to a highly expressed bacterial protein has proven particularly effective for improving the expression level of GPCRs in *E.coli*[39]. Several GPCRs have been functionally expressed in *E.coli* as fusion proteins[40-42] and refolded like human leukotriene B4 (LTB4) receptor BLT1 from

inclusion bodies[43]. GPCR expression has been achieved at relatively high levels in *S.cerevisiae* and *S.pombe* but *P.pastoris* generated the best expression results. Insect and mammalian expression is extremely expensive from a NMR point of view even though it serves as the optimal host for GPCR expression. Thus generally insect and mammalian cell expression systems have been most frequently employed for pharmacological development. For biophysical and structural studies that require large amounts of GPCRs, yeast and *E.coli* systems are attractive for their ease of large-scale production but only with a varying degree of success.

Vector	Advantages	Disadvantages	Example	Ref
E.coli (Native)	rapid , inexpensive , scalable.	toxic to cells , no post-translational modifications fusion protein required .	human A_{2A} AR rat neurotensin receptor	[40] [44]
E.coli (Inclusion body)	rapid , inexpensive , scalable, high expression	no post translational modifications, requires refolding	BLT1	[43]
Sacchromyces cervisiae	high biomass, relatively easy to use, scalable	difficulty in clone selection, hyperglycosylation	yeast ste2R dopamine D1A	[45] [46]
Pichia pastoris	high yields	difficulty in cloning selection, thick cell wall	human β2-AR	[47]
Baculovirus	mammalian-like glycosylation	more expensive, slow virus stock production	neurokinin -1, human β2-AR	[48] [49]
Mammalian cell lines	properly folded protein	very expensive , low yield	A_{2A} AR	[50]
Cell – free translation	simple , fast	very low yields	human β2- AR	[51]

Table 2: Comparison of expression systems for GPCR biosynthesis[52].(Modified)

2.6 Current state of structural studies of GPCRs

The path to the three dimensional structure of rhodopsin was long and difficult. Up to now the molecular structures of only 5 unique GPCRs have been determined including bovine rhodopsin with and without the retinal ligand as well as opsin with a C-terminal 11-residue peptide fragment of a Gα-protein (Gα-CT)[20, 23, 24], a highly engineered human β2-adrenergic receptor with a replaced intracellular loop3 (C-III)[5, 6], and a turkey β1-adrenergic receptor with the C-III loop partly removed and most of the C-terminus deleted[7] and human A_{2A}-Adenosine receptor with several thermo stabilizing mutations and C-terminal deletion[8]. Knowledge about the three-dimensional structure of these fascinating molecules are crucial for the understanding

of its function and are of great interest from the perspective of membrane-protein biophysics, cell biology, physiology and drug discovery. X-ray crystallography, electron microscopy or diffraction, NMR spectroscopy and molecular modeling are the four techniques to obtain such structural information.

2.6.1 General structure

As shown in Figure 1 all GPCRs have an extracellular N-terminal segment, seven TMs, which form the TM core, three extracellular loops, three cytoplasmic loops, and a C-terminal segment. A fourth cytoplasmic loop is formed when the C-terminal segment is palmitoylated at Cys. Each of the seven TMs is generally composed of 20–27 amino acids. On the other hand, N-terminal segments (7–595 amino acids), loops (5–230 amino acids), and C-terminal segments (12–359 amino acids) vary in size, an indication of their diverse structures. The extracellular parts of the receptor can be glycosylated. These extracellular loops may also contain two highly-conserved cysteine residues that form disulfide bonds to stabilize the receptor structure.

2.6.1.1 Why Seven TMs?

The ubiquitous adoption of a seven TM structure raises the inevitable question concerning its structural and functional merits. Odd numbers of TMs place the N- and C-terminal segments at opposite membrane surfaces. It allows glycosylation and ligand binding at the N-terminal segment, and phosphorylation and palmitoylation at the C-terminal segment for desensitization[53] and internalization. Seven TMs may be the minimum necessary to form six loops and a TM core with a sufficient size and versatility to offer a prodigious number of specificities required for the regulatory mechanism. TM α-helices vary in length and can extend beyond the lipid bilayer. Therefore, the boundaries between TMs and loops are likely to be uneven and may be dynamic. In many GPCRs the transmembrane domains 1, 4 and 7 are significantly more hydrophobic than TMs 2, 3, 5 and 6 that contain several ionic and/or neutral residues. The seven TMs of animal rhodopsin[54] and adrenergic receptors[55] are arranged as a closed loop in the counterclockwise direction from TM 1 to TM 7 when viewed from the extracellular surface. This arrangement makes the more hydrophobic TMs 1, 4 and 7 exposed more to the lipid bilayer than the less hydrophobic 2, 3, 5 and 6. The TM core is tightly packed by hydrogen bonds and salt bridges and does not have a vacant

pocket, channel, or tunnel structure. The two conserved Cys residues in extracellular loops 1 and 2 are known to be linked by a disulfide bond in bovine rhodopsin, the Thyroid stimulating hormone (TSH) hormone receptor, the thromboxane receptor, and the Gonadotropin releasing hormone (GnRH) receptor. Thus GPCRs show a far greater conservation with regard to the three-dimensional structure than to the primary sequence.

2.6.2 The Rhodopsin structure

Rhodopsin, a classic G-protein coupled receptor also known as visual purple, is a pigment of the retina and consists of the protein moiety opsin (40 kDa) and a reversibly covalently bound cofactor, 11-*cis* retinal (a derivative of vitamin A) through a lysine residue[56, 57]. Rhodopsin is involved in the molecular transformation of light energy into a neuronal signal transmitted to the secondary neurons of the retina and ultimately to the brain. Absorption of a photon by the 11-*cis*-retinal causes its isomerization to all-*trans*-retinal[58], leading to a conformational change of the protein surface and making it capable of binding and activating G-protein which leads to the visual signaling cascade.

The crystal structure of rhodopsin has provided the first three-dimensional molecular model for a G-protein-coupled receptor (GPCR) and represented a monumental step in this field[20]. The initial structure has been refined at 2.8 Å resolution[59] (Figure 4a). The N-terminus of the protein is located on the extracellular side of the membrane and it contains five distorted strands. Two oligosaccharide sites are located at residues 2 and 15 and they are not involved in any other interactions. $(Man)_3(GluN)_{10}$ was the main glycosylation pattern found at these sites. The loops E-I and E-III run along the periphery of the molecule while a part of E-II loop folds deeply into the center of rhodopsin. Cys187 forms a disulphide bond with $Cys110^{3.24}$ at the extracellular end of helix-III.

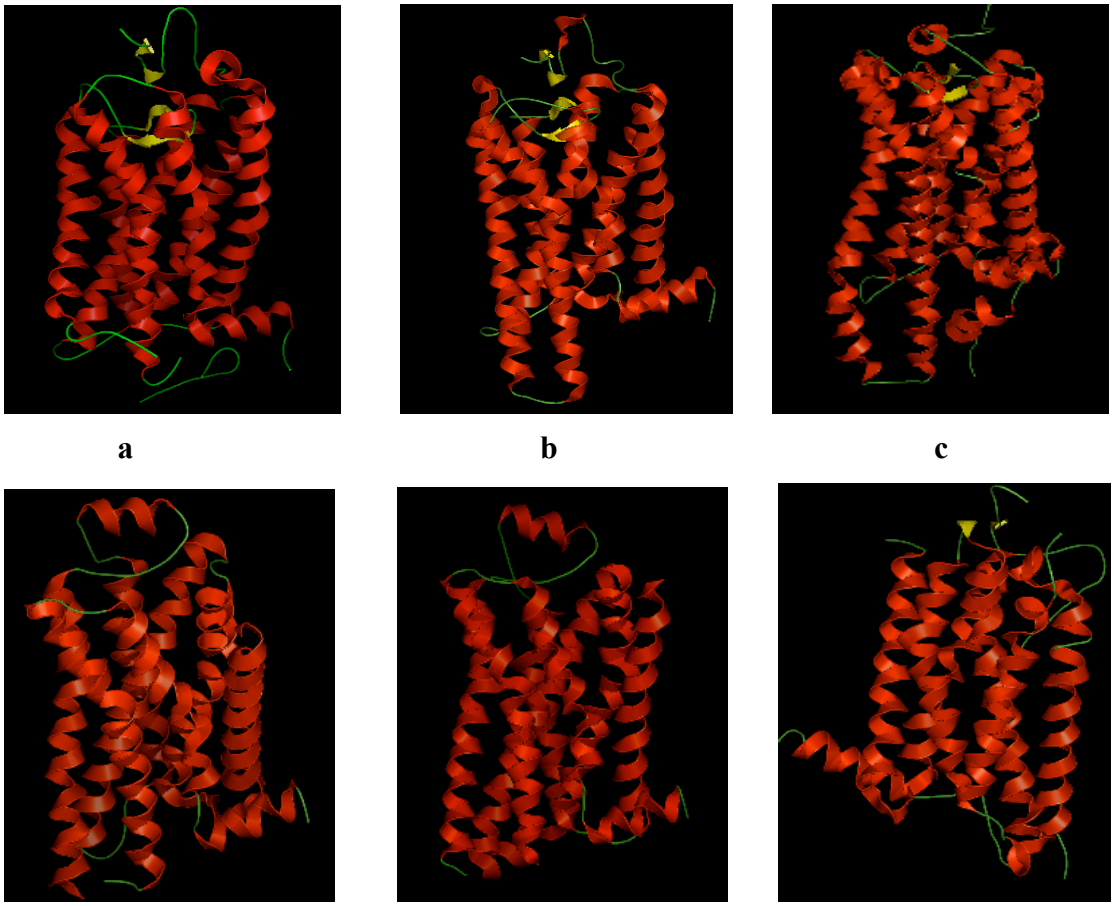


Figure 4:. Available high-resolution structures of GPCRs. (a) bovine rhodopsin; (b) bovine opsin; (c) squid rhodopsin; (d) human β2-adrenergic receptor; (e) turkey β1-adrenergic receptor,(f) human A_{2A}Adenosine receptor

2.6.2.1 Transmembrane region

The helices are irregular in length and orientation. Helix III is the longest and passes through the center of the protein. Helix VIII is a short helical segment on the cytoplasmic side of the membrane oriented with its helical axis parallel to the membrane surface. The transmembrane helices are not straight, regular α helices but kinked and bent. Proline residues are associated with the kinks except in the case for the kink in helix II at $Gly89^{2.56}$ and $Gly90^{2.57}$. In addition, it has been noted that turns of 3_{10} and π helices are also found in this structure. Two short β strands are located on the extracellular side of the protein near the retinal binding site.

2.6.2.2 Retinal binding site

11-*cis*-retinal is covalently linked to Lys296$^{7.43}$ in helix VII by forming a Schiff base linkage with the amine group of the lysine residue. The retinal is located between the helices towards the extracellular side of the protein (Figure5) and is completely buried within the protein. Towards the extracellular surface are found the two short β strands and extracellular surface loops that make up the TM plug shielding the retinal from aqueous environment. The transmembrane helices block ready access of the chromophore to the hydrophobic region of the bilayer. The chromophore is not planar. The conjugated system is twisted, possibly due to packing interactions with the protein, or due to the steric interactions within the chromophore.

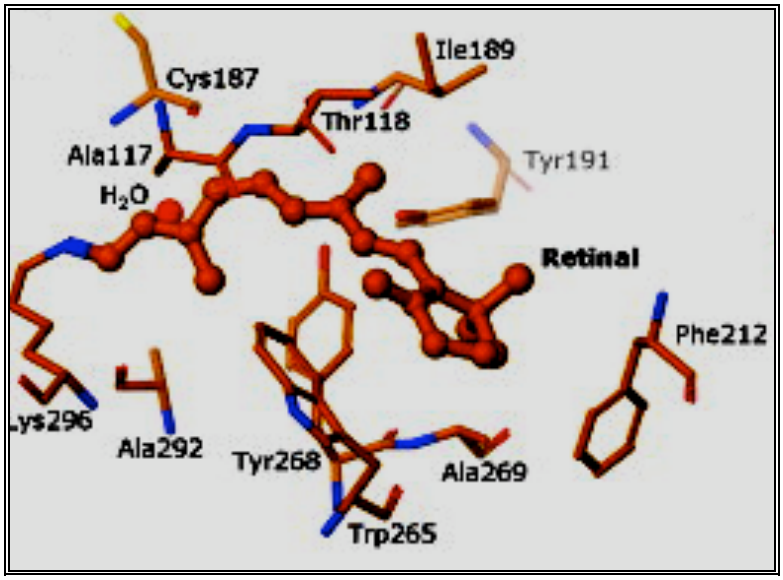


Figure 5: Binding sites of bovine rhodopsin with retinal (1U19)[60]

The chromophore binding environment is made up of a mix of hydrophobic and polar/charged groups. At one end 11-*cis* retinal, the chromophore,covalently binds to Lys296$^{7.43}$.and at the other end, its β-ionone ring is buried in a hydrophobic pocket formed by Trp265$^{6.48}$, Phe212$^{6.47}$, and Tyr268$^{6.51}$. The interaction between the β-ionone ring and Trp265$^{6.48}$ forces the side-chain rotamer conformation of Trp265$^{6.48}$ to be that of the inactive state. Switching between this and the active Trp265$^{6.48}$ conformation initiates the so-called ‘‘toggle switch’’ for activation/inactivation of rhodopsin. Polar groups such as Thr118$^{3.33}$ and Tyr268$^{6.51}$ are located near the center of the polyene, and the side chain of Glu122$^{3.37}$ is close to the β-ionone ring. Also, Glu183 makes a close water-mediated approach to the retinal. The Schiff base at the other end of the

chromophore is protonated in the activated form of the protein, and Glu113$^{3.28}$ serves as a counterion for it in the interior of the protein.

2.6.2.3 Cytoplasmic surface

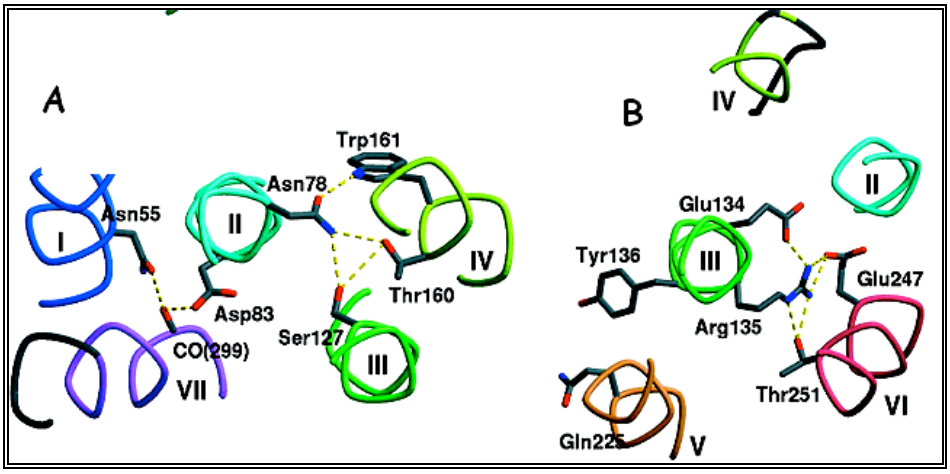
The C-I loop exhibit rigid organization. The C-II loop exhibits an L-shaped structure with a beta barrel (Met143 to Phe146) almost along the main axis of rhodopsin. The C-III loop is highly flexible and it should be noted that its length varies considerably in different GPCRs and may be critical for the functionality and specificity in the G-protein activation. The C-IV loop in the cytoplasmic region is nearly perpendicular to H-VII and it follows the NPXXY motif a highly conserved region of GPCR. The side chains of Cys322 and Cys323 point to the outside of rhodopsin indicating its probable palmitic acid attachment site.

2.6.2.4 Intramolecular interactions and GPCR activation

Ground-state rhodopsin is inactive, and several of its structural elements combine to restrain the structure. Interactions among the extracellular loops, including a disulfide bridge limit the conformational flexibility of this part of the molecule under dark, non-signaling conditions. In addition, the interactions between the chromophore and the protein, both hydrophobic and electrostatic, tighten the inactive receptor structure. Mutations of Lys296$^{7.43}$ or Glu113$^{3.28}$ eliminating these interactions result in constitutively active receptors. The transmembrane region of rhodopsin is stabilized by a number of interhelical hydrogen bonds and hydrophobic interactions, and most of them are mediated by highly conserved residues in GPCRs. One of the residues that exhibit the highest conservation is Asn55$^{1.50}$ in H-I. Its side chain is responsible for two interhelical hydrogen bonds to Asp83 in H-II and to the peptide carbonyl of Ala299$^{7.46}$. Asp83$^{2.54}$ is in turn connected via a water molecule to the peptide carbonyl of Gly120$^{3.35}$ in H-III (Figure-6A). Another region that mediates constraints for these helices includes Asn78$^{2.49}$ of H-II, which is hydrogen-bonded to OH groups of Ser127$^{3.42}$ of H-III and Thr160, Trp161 of H-IV. Helices H-III, H-IV and H-V can be also linked through interaction among Glu122$^{3.37}$, Met163, and His211$^{5.46}$ (Figure 6A)

The tripeptide Glu134$^{3.49}$, Arg135$^{3.50}$, Tyr136$^{3.51}$ is part of a highly conserved (D/E)R(Y/W) motif found in GPCRs. These residues participate in several hydrogen

bonds with surrounding residues and form an "ionic lock" which is responsible for holding the receptor in its inactive state (Figure 6B). The carboxylate of Glu-134 forms a salt-bridge with the guanido moiety of the next Arg-135$^{3.50}$. Arg135$^{3.50}$ is also connected to Glu247$^{6.30}$ and Thr251$^{6.34}$ in H-VI. Val137$^{3.52}$, Val138$^{3.53}$, to Val139$^{3.54}$ are also closely located to partly cover the cytoplasmic side of Glu134$^{3.49}$ and Arg135$^{3.50}$. These could be one of the critical constraints keeping rhodopsin in the inactive conformation.



Structural details of the intramolecular interactions in Rhodopsin [20]

Figure6A: Interhelical hydrogen bonds mediated by a highly conserved Asn55, connecting H-I, H-II, and H-VII, and by Asn78 for H-II, H-III, and H-IV.

Figure6B: The tripeptide region, Glu134$^{3.49}$, Arg135$^{3.50}$, Tyr136$^{3.51}$ known as a (D/E)R(Y/W) motif located near the cytoplasmic end of H-III forming an ionic lock with residues in H-VI.

The first step in the activation of rhodopsin is the photoisomerization of 11-*cis*-retinylidene into all -*trans*-retinylidene after absorption of a photon. This isomerization is fast, and slower conformational adjustments in the protein as well as the chromophore eventually give rise to the signaling form of the protein, metarhodopsin II. The conformational switching of the chromophore might require reorganization of the helical structure of helix VII which disrupts the salt bridge between Glu113$^{3.28}$ and Lys296$^{7.43}$ and movement of a proton. In addition to this, the helix VII moves away from helix I and helix VI moves away from the other helices. Glu247$^{6.30}$ is no longer able to interact with Arg135$^{3.50}$, which is then able to reorient to the cytoplasmic

surface, where it can interact with transducin. Protonation of Glu134$^{3.49}$ also occurs during this conformational transition. Motion of the helix VII has been identified as important for the interaction with G-protein. The major structural features of rhodopsin structure are represented in the Figure 7.

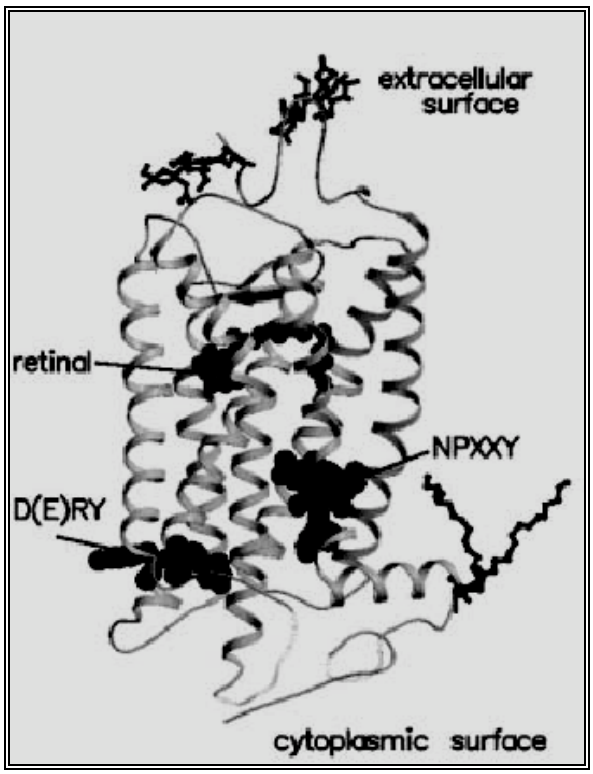


Figure 7:. Stereoview showing the location of structural features in rhodopsin[61]. The polypeptide backbone is shown as gray ribbon. Retinal is shown as small black spheres. Residues in the D(E)RY and NPXXY motifs are shown as big black spheres.

2.6.3 Opsin/Rhodopsin in the activated state

Comparison of the rhodopsin structure with invertebrate squid rhodopsin (Figure 4c), which couples to Gq, reveals that the most distinguishing feature of these structures occurred in the intracellular region of helix-V and helix-VI and the intervening loop C-III that has a 12-residue insertion compared with rhodopsin. In this region, these two helices are longer and rigid, extending well away from the core,[62] and form one side of a binding region for an occluded octylglucoside, which indicates that the surface around this extended region may be needed to bind the G-protein.

Two bovine opsin papers appeared recently last year. In the more recent one opsin is complexed with an eleven amino acid peptide, GαCT, derived from the C-terminal of the transducin $G\alpha_t$ protein (Figure 4b). Opsin is the retinal free photoreceptor protein generated after photoactivation and Schiff base hydrolysis of rhodopsin. In these structures more dramatic structural changes are observed at the cytoplasmic surface especially in the C-III loop and intracellular region of helix-V and helix-VI. Residues of the highly conserved E(D)RY and NPXXY motifs play key roles in these changes. The end result of the changes from inactive rhodopsin to active-state opsin is the creation of a cavity between helix-III, helix-V and helix-VI, in which the G protein transducin can bind.

2.6.4 Adrenergic receptors

The adrenergic receptors (or adrenoceptors) belongs to class A GPCRs that are targets of the catecholamines, especially noradrenaline (norepinephrine) and adrenaline (epinephrine). The adrenergic receptors in the amine group of class A GPCRs consist of two main subfamilies, α and β, which differ in tissue localization and ligand specificity as well as in G protein coupling and downstream effector mechanisms[63]. α-receptors have the subtypes α1 (a G_q coupled receptor) and α2 (a G_i coupled receptor) and β receptors have the subtypes β1, β_2 and β3. All three are linked to G_s proteins (although β2 also couples to G_i)[64], which in turn are linked to adenylate cyclase. Genetic modifications of adrenergic receptors are associated with diseases as diverse as asthma, hypertension, and heart failure[65].

2.6.4.1 β_2-adrenergic G-protein coupled receptor

The high-resolution β2-AR X-ray structure,(figure 4d) bound to the inverse agonist carazolol and timolol was published recently[5, 21, 22] . The approach of splicing T4L into the C-III region and complexation of β2-AR with Fab portion of the monoclonal antibody (Mab5) raised against the C-III region of β2-AR were utilized to provide conformational stability. The binding of timolol to β2-AR was very similar to that of carazolol. In contrast to rhodopsin, instead of short β-sheet the E-III loop regions of β2-AR contain a short α-helix that was stabilized by intra- and inter-loop disulphide bonds, and the cytoplasmic N-terminal regions were disordered. The overall positions of the ligands are slightly more extracellular than 11-*cis* retinal in rhodopsin.

This leads to significant difference in inverse agonist–antagonist interactions with the residues Trp286$^{6.48}$ (human β2-AR) and Trp303$^{6.58}$(avian β1-AR), which were suggested to undergo key rotamer conformational transitions in GPCR activation, referred to as the 'rotamer toggle switch[66]. Ligand binding sites of different GPCRs were compared in Figure 8 in the form of 2D interaction map.

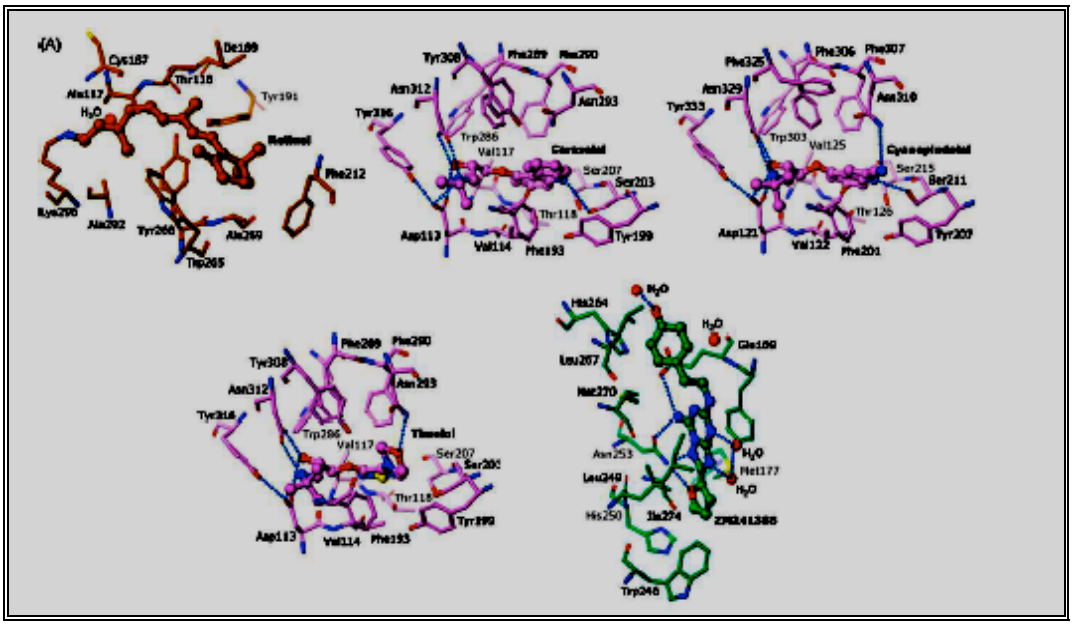


Figure 8: Binding sites of bovine rhodopsin with retinal (1U19); the human β2-adrenergic receptor with carazolol (2RH1); the turkey β1-adrenergic receptor with cyanopindolol (2VT4); the human β2-adrenergic receptor with timolol (3D4S); and the human A_{2A} adenosine receptor with ZM241385 (3EML)[60].

The 'ionic lock' between the highly conserved E/DRY motif on TM3 and a glutamate residue on TM6 which is seen in rhodopsin is absent in all the ligand-bound GPCR structures β_1AR, β_2AR and β_2AR. Instead polar contact between adjacent acidic and basic residues on TM3 (Glu134$^{3.49}$ followed by Arg135$^{3.50}$ of the E/DRY motif is maintained (Figure 9). However, recent MD calculations proposed that the ionic lock may also be present in these structures atleast part of the time. Another difference between the structures is at the cytoplasmic surface found in the intracellular loop 2 region (C-II), which includes a short α-helix in the β1-AR and A_{2A}AR structures that is absent in β2-AR and rhodopsin. This structure serves as a platform for a hydrogen-bonding interaction of a conserved tyrosine (on C-II) with the E/DRY motif (on TM3); its absence in β2-AR could help to explain the higher relative basal activity of this

receptor. In the timolol-β2-AR structure two cholesterol binding sites, which are not involved in crystal packing, are revealed. These binding sites may play a role in cholesterol-mediated thermal stabilization, allosteric modulation of ligand binding to the high-affinity agonist binding state, and receptor trafficking (see Ref. [22] and references therein).

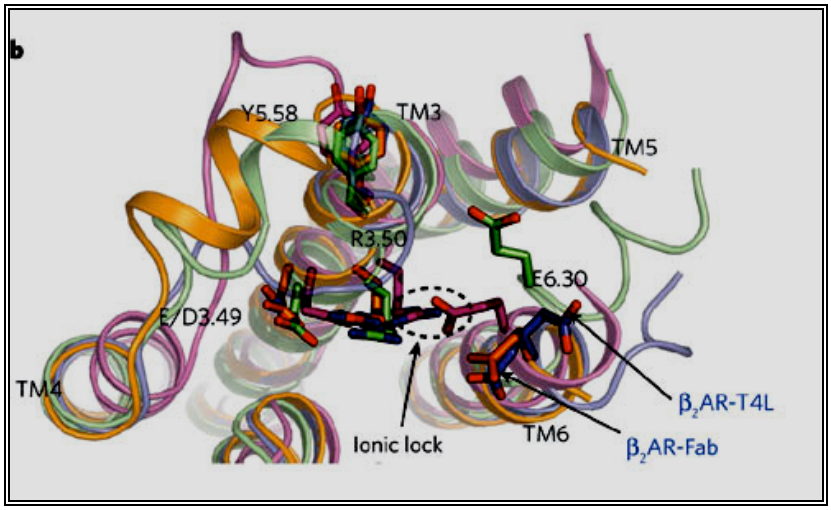


Figure 9: The ionic-lock residues at the cytoplasmic end of TM3 ($R^{3.50}$ and $E/D^{3.49}$), and TM6 (E6.30) for the four superimposed receptor structures bovine rhodopsin (purple, bound to 11-*cis* retinal), avian β1AR (orange, bound to cyanopindolol), human A2A adenosine receptor (green, bound to ZM241385) and human β2AR (blue, bound to carazolol) are shown. $R^{3.50}$ engages $Y^{5.58}$ on TM5, rather than $E^{6.30}$ on TM6 in the opsin 'active state'. The rotameric position of $E^{6.30}$ differs for the two β2-AR structures[67].

2.6.4.2 β1-adrenergic G-protein coupled receptor

Another GPCR structure was from the turkey β1-adrenergic receptor bound to the cyanopindolol[7]. Because of the difficulties in crystallization of human β1-AR the studies were done with turkey (*Meleagris gallopavo*) β1-AR. Through extensive analyses of the thermal stabilizing effects of various mutants and their combinations, a composite of six mutations was introduced to sufficiently stabilize the complex without the introduction of a companion (mAb or T4L) protein. There were truncations of some residues from N and C-termini as well as from the C-III loop. The close relationship of these proteins, i.e., β1-AR *vs*. β2-AR, makes the comparison of structures particularly meaningful. At the sites of the six point mutations, there was no evidence for any significant changes in backbone conformation.

The overall structure was found to be very similar to β2-AR. (Figure 4e). The structures of the three extra-cellular loops in β1-AR are very similar to those of β2-AR. In case of cytoplasmic loops, the C-I loop was similar for rhodopsin, β1-AR and β2-AR, but there are major differences in C-II and C-III. The third cytoplasmic loop cannot be compared as it was subjected to several modifications in all the GPCR structures except rhodopsin. In β1-AR, C-II forms a short α-helix (Pro146$^{3.57}$ to Leu152$^{3.63}$) parallel to the membrane surface whereas in both β2-AR structures and in rhodopsin it is in an extended conformation. Among the more notable differences found between the β1-AR and β2-AR structures are the residues in the ‘‘ionic-lock’’ region. The ‘‘ionic lock’’ observed in rhodopsin was not formed, indicating that that the ionic lock was not an essential feature of inactive state of GPCRs. But the short α-helix which was formed in C-II forms an hydrogen bond between Tyr149 on C-II and Asp138$^{3.49}$ of the ‘‘ionic lock’’ DRY motif on H-III which would explain the inactivity of the cyanopindolol β_1AR structure in contrast to the residual basal activity of the two β_2AR complexes with inverse agonists.

The mode of binding of cyanopindolol to β1-AR was found to be similar to that of carazolol in β2-AR(Figure 8).. However, the extra ring in the carazolol heterocyclic ring, owing to van der Waals contacts with Tyr199$^{5.38}$ in β2-AR pushes the ligand more deeply into the binding site. The nitrogen in the cyano-moiety of cyanopindolol makes a weak hydrogen bond with the hydroxyl of Thr203$^{5.34}$, which was located together with Phe201$^{5.32}$ on E–II. To account for the subtype specificity and the fact that some ligands preferentially bind to either β1-AR or β2-AR, there must be differences in amino acid residues close to the ligand-binding pocket that directly or indirectly affect binding. In between these two adrenergic receptors the respective residues were found to be Val172$^{4.56}$ and Phe325$^{7.35}$ in β_1AR which was equivalent to Thr164$^{4.56}$ and Tyr308$^{7.35}$ in β_2AR. These differences introduce polar residues near the binding pocket of β1-AR relative to β2-AR, which could affect ligand selectivity.

2.6.4.3 Architecture of the Human A_{2A} Adenosine receptor

A recent publication has now provided the X-ray structure of the A_{2A} adenosine receptor in complex with the high-affinity antagonist ZM241385[8] (Figure 4f) .In

order to overcome thermal instability problems, human A_{2A}Adenosine receptor was expressed as T_4-Lysozyme (T_4L) fusion where most of the C-III loop (Leu209$^{5.70}$ to Ala221$^{6.23}$) was replaced with T_4L and the carboxy terminal tail (Ala317 to Ser412) was removed to improve crystallization. In addition the N-Linked glycan associated with Asn154$^{4.73}$ has been removed enzymatically. The structure provides yet more information and new insights about GPCRs. There are three major features, which are different from the previously reported GPCRs. In the A_{2A} AR structure, the organization of the extra cellular loop is markedly different from the other GPCR structures reported. It resembles neither the extended β-sheet of rhodopsin nor does it include the α-helix structure of the adrenergic receptors,rather, E-II adopts a random coil conformation that has three cysteine bridges to E-I and one within E-II, resulting in an opening that could allow the entry of small molecules into the active site.

Secondly, the antagonist ZM241385 binds in an extended conformation perpendicular to the plane of the membrane and colinear with transmembrane helix VII, (Figure 8) while interacting with both ECL2 and ECL3. The bicyclic triazolotriazine core of ZM241385 is anchored by an aromatic stacking interaction with Phe168$^{5.29}$, an aliphatic hydrophobic interaction with Ile274$^{7.39}$ and a hydrogen-bonding interaction with Asn253$^{6.55}$ (Figure 6). Adjacent to Phe168$^{5.29}$ the polar residue Glu169$^{5.30}$ interacts with the exocyclic amino group (N15 atom) linked to the bicyclic core of ZM241385. Finally the seven transmembrane arrangement was similar to the other GPCRs, however there was shift in the relative orientation of the various helices. The conserved tryptophan residue on helix VI known as the toggle switch interacts with the ligand and modulates the receptor equilibrium between active and inactive state. An interaction between the D/ERY motif and Helix VI, which constitutes an ionic lock that restrains the fully inactive conformation of rhodopsin, was not seen in the A_{2A} AR. Instead of participating in an ionic lock, the arginine residue in the D/ERY motif may play a role in stabilizing the deprotonated state of the adjacent aspartate or glutamate residue, which would strengthen the polar interactions between the D/ERY motif and both C-II and helix II. This set of interactions may have direct implications in G-protein activation[68]. The A_{2A} AR ligand–bound structure suggests that there was no general, family-conserved receptor-binding pocket, in which selectivity was achieved through

different amino acid side chains. Rather, the pocket itself can vary in position and orientation so as to offer more opportunity for receptor diversity and ligand selectivity.

2.6.5 Comparative analysis of GPCR crystal structures

The GPCRs share many hallmark motifs and structural homologies eventhough their primary sequences are quite divergent .Structures of the ligand mediated β1-AR, β2-AR, and the A_{2A} adenosine GPCRs are providing significant advances in the understanding of the partially/fully inactive states, albeit more clarity is needed. These structures also provide a valuable template for developing homology models for agonist binding and for understanding of the mechanism of GPCR activation. Comparison of the opsin/rhodopsin GPCR structure is again leading the way in understanding the mechanisms of activation at the atomic level.

Summary of structural modifications required to obtain crystals of G-protein-coupled receptors (GPCRs)

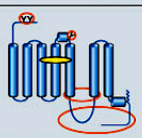
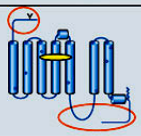
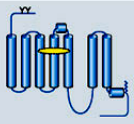
GPCR	Topology	Ligand (indicated by yellow oval)	Method of stabilization	Other modifications (indicated by red ovals and circles)
β_2AR-Fab		Inverse agonist for stabilization	Stabilized TM5/TM6 (transmembrane) region through binding to Fab, which improves crystal lattice-forming contacts	Truncated flexible C terminus (potential Ser/Thr phosphorylation sites) Removed N-linked glycosylation N-terminal affinity tag
β_2AR-T4 lysozyme Adenosine A_{2A}-T4 lysozyme		Inverse agonist or antagonist for stabilization	Stabilized TM5/TM6 region through T4L insertion, which improves crystal lattice-forming contacts	Truncated flexible ICL3 (intracellular loop) and C terminus (potential Ser/Thr phosphorylation sites) Removed N-linked glycosylation N- or C- terminal affinity tag
β_1AR		Antagonist for stabilization	Mutations to increase thermal stability and functional expression (indicated by blue circles)	Truncated flexible ICL3, N- and C-termini (glycosylation and Ser/Thr phosphorylation sites) Removed palmitoylation site C-terminal affinity tag
Rhodopsin/opsin		Crystal structures obtained with and without bound retinal	Native protein crystallizable without modifications	Not required

Table3: Challenges in GPCR crystallography[67]

At the same time one must be concerned about the modifications that facilitated crystal formation could produce high resolution structure which may not completely the resemble the WT receptors since the fusion protein exhibited slightly elevated agonist binding affinities; however, antagonist binding affinities were normal, and only minor differences were observed. A summary of the structural modifications required to obtain crystals of GPCRs are outlined in the Table 3. Using fluorescent probes, the authors showed that agonists can induce protein conformational changes consistent

with receptor activation. Nevertheless, a full understanding of the structural basis of GPCR activation will require a high-resolution structure of a complex between a receptor with an agonist bound and its G-protein, as well as methods such as NMR to assess dynamics of their interaction that cannot be captured by the static snapshots provided by x-ray crystallography. The important features of the known GPCR structures and their significance are presented in the Table 4 and 5

GPCR type	**Characteristic Structural determinants**	**Functional interpretations**
Rhodopsin + 11-*cis* retinal -- 2.8 Å	ECR- one disulphide bridge. 4 β sheets interact with N-terminus and C-II loop forming compact retinal plug. 11 – *cis* retinal interacts with F212 (TM5),Y268and W265(TM6),TM3 and C-II loop. Ionic lock between E134 ,R135 (DRY motif ;TM3) and E247T251 (TM6) constraining the TM6 helix bundle inside . Y306 from the NPxxY motif (TM7) connected to F313 from cytoplasmic helix H8.	Ionic lock and absence of rotamer toggle switch – responsible for absence of basal activity (inactive atate). Retinal plug seals retinal binding pocket .
Squid rhodopsin + 11-cis retinal 3.7 Å	ECR- one disulphide bridge. TMR: longer TM5 and TM6 with hydrophilic extension in cytoplasm forming a compact structure. ICR: two cytoplasmic helices H-VIII and H-IX linked by a hydrophobic 3_{10} helix, with H-IX interacting with C-II, TM5, TM6 and H-VIII.	The TMs protusion from the membrane surface may specify the recognition and the coupling mode with $G\alpha_q$ proteins.
Bovine ligand free opsin in its active G-protein interacting conformation No ligand –3.*2* Å	Dimers connected via TM1 and H-VIII. ECR: one disulfide bridge, 'retinal plug' structure is maintained. Two openings in the retinal binding pocket between TM5/TM6 and between TM1/TM7. The 'ionic lock' is broken and two new interactions (R135-Y223 (TM5), E247/T251-K231 (TM5)) are formed that lock and stabilize TM5/TM6 pairing. This provides crevice for GαCT interactions and binding. ICR: GαCT connects DRY (R135, TM3) to NPxxY (N310, TM7) and H-VIII (Q312) via a hydrogen bond network.	Two openings – gates for one selective for 11-*cis* retinal and other for 11-*trans* retinal . Retinal plug seals retinal binding pocket. Insights into the structural changes involved in signal transfer from the receptor to G-protein.

Table 4 : Comparative analysis of the known rhodopsin structures [69](Modified).

Abbrevations: ECR:Extracellular region, ECL: Extracellular loop, TMR: Transmembrane region, TM1-7: Transmembrane helices, ICR: Intracellular region , ICL: Intracellular loo

GPCR type	Characteristic Structural determinants	Functional interpretations
β2AR – Fab + carazolol – 3.4 Å /3.7 Å	ECR: not visible. TMR: absence of the 'ionic lock' (R131 is too far from E268) Weaker interactions between TM3 (DRY) and TM6 leading to a more open structure.	The more open structure may account for the high basal activity and the structural instability.
β2AR – T4 lysozyme + carazolol - 2.4 Å	ECR: N-terminus not visible; unusual pair of disulfide bridges and a short helical segment not present in rhodopsin (E-II). Fusion with T4-lysozyme in place of C-III provides conformational stability and increases polar surface). TMR: presence of 3 molecules of cholesterol per monomer Absence of the 'ionic lock' (R131 is too far from E268).	This structure provides details on the interactions with carazolol.
β2AR – T4 lysozyme +timolol 2.8 Å	Specific cholesterol binding site between TMs 1, 2, 3 and 4 (2 molecules of cholesterol per monomer) . Compared to carazolol, timolol binds deeper into the receptor.	A consensus motif predicts cholesterol binding for 44 % of class A receptors . Cholesterol enhances thermal stability and shifts the receptor to the high affinity agonist binding state. Allosteric modulation by cholesterol .
Mutated turkey ß1AR + cyanopindolol - 2.7 Å	ECR: two disulfide bridges stabilizing an α-helix in ECL2 TMR: 15 residues (TM3, 5, 6, 7 and ECL2) make contact with cyanopindolol. Hydrogen bond between DRY (TM3) and Y149 (ICL2) In the binding pocket, V172 and F325 could affect ligand selectivity. Absence of the 'ionic lock'. ICR: short helical segment in C-II parallel to the membrane.	Structured C-II and its related interactions may explain receptor low basal activity. Both Y149 and the ICL2 α-helix have key roles in G-protein coupling. In the binding pocket, V172 and F325 could affect ligand selectivity.
Adenosine A2A receptor –T4 Lysosome + ZM241385 - 2.6 Å	ECR: four disulfide bridges constraining random coiled ECL2. Absence of the 'ionic lock'. TMR: the binding pocket is shifted closer to TM6 and TM7 ZM241385 interacts with the 'toggle switch' W246 Hydrogen bonding between D101 (TM3) and Y112 (ECL2) ICR: h8 along the membrane interacting with TM1 Short helical segment in C-II.	Extensive disulfide bond network forms a rigid, open structure which forces the antagonist to bind perpendicularly to the membrane ZM241385 restricts movement of W246 (involved in the receptor activation) - stabilize receptor inactive state. DRY motif participates in the interactions that restrain ICL2 conformation and explain receptor low basal activity.

Table 5: Comparative analysis of the known GPCR structures [69](Modified).

Abbrevations: ECR:Extracellular region, ECL: Extracellular loop, TMR: Transmembrane region, TM1-7: Transmembrane helices, ICR: Intracellular region , ICL: Intracellular loop

2.7 GPCR oligomerization

Accumulating evidence of GPCR oligomerization on the one hand and perfect functionality of monomeric receptors on the other creates an impression of controversy. This is an actively-studied area in GPCR research. Do GPCRs function as monomers, dimers or higher-order oligomers? It is very difficult to provide a solid answer for this question. Different GPCR subtypes, and even the same receptor at different stages of its life cycle, most likely exist in different oligomerization states, from monomers to dimers and possibly higher-order oligomers. Another important issue is the stability of the oligomers, which might be permanent (as seen in Family 3 GPCRs) or very transient or anything in between. One more important aspect that has to be kept in mind is that several types of GPCRs reside in micro domains covering a small fraction of the cell surface[70], [71, 72]. Crowding of over expressed receptors in these micro domains is sufficient to yield a FRET signal without oligomerization[70].

Another problem concerning the monomer–dimer debate is the unambiguous interpretation of inherently ambiguous data. Therefore, we should avoid overgeneralization. An amazing structural and functional diversity of the GPCR superfamily acquired in more than a billion years of evolution makes it highly unlikely that every receptor performs every function in exactly the same state. The best-studied example of receptor oligomerisation are the metabotropic $GABA_B$ receptors[73]. These receptors are formed by heterodimerization of $GABA_BR1$ and $GABA_BR2$ subunits. Only expression of the two subunits together leads to plasma membrane expression of functional receptor[74, 75].

2.7.1 GPCR studies in fragments

Despite fairly good methods developed for both prokaryotic and eukaryotic expression of many soluble cytoplasmic proteins, as mentioned above GPCRs have presented major challenges for obtaining high yields in almost every expression system. The low expression levels of these seven transmembrane receptors are due to the requirements for folding and transport and the cellular toxicity related to heterogenous expression in membranes. Another problem which intervenes in the GPCR structural studies by NMR is the size of the protein-detergent complex which leads to

unfavourable T_2 relaxation and broad overlapping resonances. To defeat this relaxation problem a partial solution came in the form of deutration, but this requires that backbone NHs must be back-exchanged to hydrogen. The unfolding and refolding experiments to carry out this exchange works fine with β-barrel membrane proteins but for α-helical transmembrane proteins is often problematic. Therefore a critical need arises to look for alternative approaches. One very useful method is the study of suitably chosen fragments. A significant number of studies provided structural information on fragments corresponding to sequences of GPCR helices, loops and termini. Eg: *Saccharomyces cerevisiae* pheromone receptor[76], cannabinoid receptor[77] and Neurokinin-1 receptor[78]. Significant similarities between the structures of fragments and the rhodopsin structure were observed.

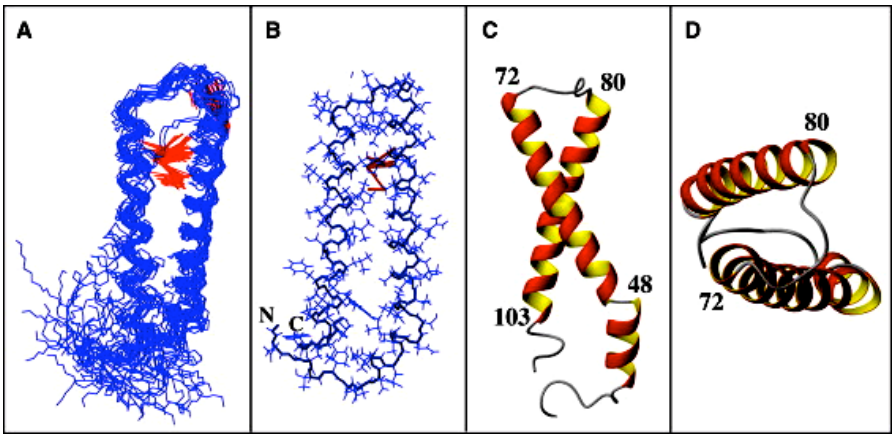


Figure10: (*A*) Backbone representation of the ensemble of the 20 lowest-energy conformers of Ste2p(G31-T110) superimposed over backbone atoms in the region comprising residues 39–103. Observed long-range NOE contacts are highlighted in red. (*B*) A single conformer from the ensemble additionally displaying the side chains. (*C*) Structure of a single conformer—view from the side of membrane interior. (*D*) The same as *C* but viewed from the cytoplasmic side.Ref:83

According to Popot and Engelman the transmembrane domains can be thought of as independent folding units and can be studied separately[79]. This hypothesis is further tested by studying individual peptides corresponding to fragments (turn or a transmembrane helix) of rhodopsin, and their structures were overlaid on the corresponding part of the crystal structure of native rhodopsin. Good agreement was observed between helices in the protein and the structures of the fragments and turns in the protein and the structures of the fragments[80-82]. From our group also Dr. Alexey

neumoin studied a construct comprising the first 2 TM domains of the yeast Ste2p receptor in LPPG micelles[83](Figure 10).

2.7.2 Helix –helix interactions

Helix-helix interactions have been discovered to contribute to stability of GPCRs. Specific interfacial interactions between transmembrane α-helices may be important in the assembly and function of integral membrane proteins. This was studied extensively for the A_{2A} receptor. These studies reported that the neighbouring helices enhance the helix stability[84] and also provide evidence for the self –association of transmembrane helix-V which could give some insights for the dimerisation of receptors[85]. The presence of motifs like GxxxG, SxxSSxxT (polar clamp) and SxxxSSxxT motifs (serine zipper)[86] have been shown to play important role in mediating helix-helix interactions and for oligomerisation of membrane proteins. These interactions are specific and strong as reflected in the ability of separately expressed bundles of GPCR transmembrane helices to associate correctly[87]. These findings suggest that some TM domains might need to interact with other domains to properly insert and fold in hydrophobic environments.

2.7.3 A membrane protein folding model

Unlike soluble proteins, intrinsic membrane proteins are designed to fold and exist in a milieu from which water is largely excluded. The lipid bilayer environment strongly limits the range of possible structures for transmembrane proteins and makes the folding, stability of integral membrane proteins easier to understand than that of their soluble counterparts. Only two structural motifs have been observed for membrane proteins (MPs), the membrane-spanning α-helix bundles and β-barrels, the former being predominant. In 1990, Popot & Engelman[79] have proposed a two stage model for helix-bundle membrane protein folding. According to this model, (Figure11) initially transmembrane (TM) helices insert into the bilayer with appropriate topology. The second stage involves side by side packing of these preformed helices to produce the final folded structure. Individual helices are regarded as autonomous folding domains.

In the first step partitioning of unfolded polypeptides into the water-membrane interface results in increased formation of backbone hydrogen bonds resulting in formation of secondary structure. Interactions of the side chains with the lipid environment lead to the insertion of transmembrane helices into the bilayer interior. Once a hydrophobic sequence has been inserted into a membrane as a helix, the net hydrophobicity of these segments precludes their dissociation from the membrane, while the cost of breaking hydrogen bonds within a low-dielectric medium prevents the helices from unfolding. The second stage is the interaction of the helices to form the tertiary fold of the polypeptide. Factors that could contribute to the energetics of stage II are the links between helices, packing of helices and lipid molecules, polar interactions between helices, and, when applicable, association with prosthetic groups or with other proteins.

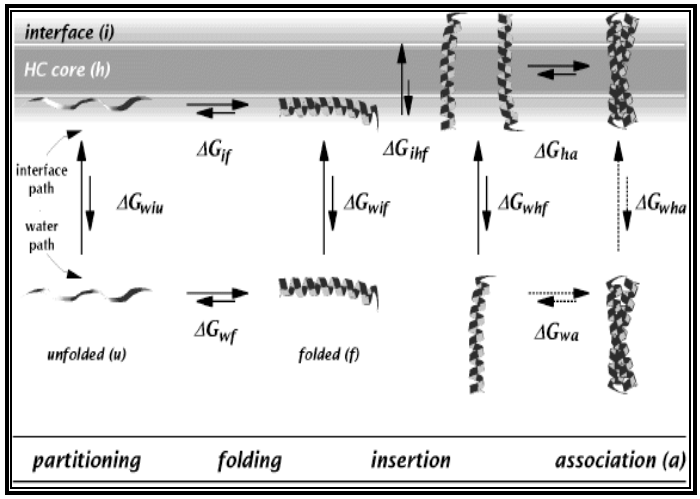


Figure 11: Steps in membrane protein folding. (From ref 79)

2.7.4 Membrane mimetics

GPCRs as integral membrane proteins can only exert their functions when they are inserted in the membrane. Their requirement for lipids has hampered structure determination by conventional approaches. The ability to perform NMR experiments on either liquid or solid-state samples provides the opportunity for structure determination of these proteins in several membrane mimetic environments. Current model membrane systems that have been employed include organic solvents, micelles, mixed micelles, bicelles, vesicles and short-chain/long-chain unilamellar vesicles (SLUVs)[76, 88, 89]. The more commonly used membrane mimetics are discussed in the following paragraphs.

2.7.4.1 Organic solvents

Mixtures of organic solvents are often used as membrane mimetics in structure determination of transmembrane proteins by solution NMR. Organic solvent mixtures can interact with the transmembrane domains of membrane proteins so that they can be solubilized without the need for any solvating lipid or detergent. The most commonly used organic solvent or solvent mixtures are DMSO,50% trifluoroethanol-water, 4:4:1chloroform:methanol:water, 80:20 hexafluoroisopropanol:water etc[81, 90]. The properties of proteins in the organic solvents /organic solvent mixture have shown to mimic those of the protein in the native environment. But most membrane proteins include significant extramembranous water-exposed domain(s), which are not expected to remain natively folded in organic solvent mixtures. Hence to summarize, organic solvent mixtures may be worth trying as membrane mimetic media for solubilizing certain membrane proteins. However, great care should be taken to establish whether or not native structure is present before spending much time on making NMR resonance assignments, which may represent misplaced effort, if the protein is not properly folded.

2.7.4.2 Micelles

A Micelle is an aggregate of surfactant molecules dispersed in a liquid colloid. When detergents are dispersed in water, they are mainly monomeric. However when the concentration increases above the critical micelle concentration (CMC), monomeric detergent molecules start associating and form a spherical aggregate termed a micelle (Figure 12). Hydrophilic head groups point outward interacting with water while the hydrophobic tails point inward forming an apolar environment. Detergents can be classified into the following categories according to charge: non-ionic (sugar-derived detergents) and ionic detergents. Among ionic detergents there are cationic, anionic (SDS) and zwitterionic (CHAPS, DPC) detergents. Detergent micelles are by far the most widely used membrane mimetic for solution NMR studies[91, 92] however, it is clear that the upper size limit for NMR structure determination of membrane proteins in micelles is considerably lower than it is for globular proteins in aqueous solution. Indeed, for membrane proteins with more than 50 residues, resonance assignments require uniform ^{13}C and ^{15}N labeling and deuteration is often also necessary.

2.7.4.3 Bicelles

Bicelles are discoidal lipid aggregates composed of long-chain phospholipid and either detergent or short-chain phospholipids (Figure12). The distinguishing structural feature of a bicelle is a central planar bilayer formed by the long-chain phospholipid, surrounded by a rim of short-chain phospholipids or detergent that shields the long-chain lipid tails from water. Bicelles offer the opportunity to study membrane proteins in an environment that closely resembles the planar surface of biological membranes[88]. The bicelle size is controlled both by the ratio of long-chain:short-chain phospholipids and the total phospholipid concentration. Bicelle discs orient in the magnetic field; therefore, it is possible to measure parameters containing long-range angular information, such as residual dipolar couplings. Since the long-chain phospholipid in bicelles is sequestered into the planar core region, devoid of short-chain phospholipid or detergent, the core region of the bicelles mimics a section of natural membrane much better than standard micelles. Bicelles have proven to be very useful in solid-state NMR applications, and high-quality spectra are often observed due in part to the highly liquid crystalline environment of the bicelle lipids.

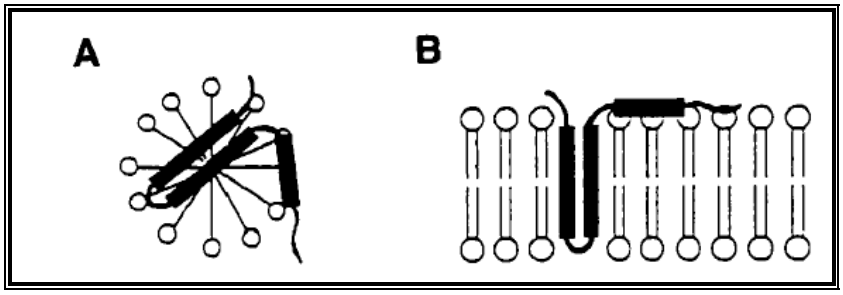


Figure 12: Representation of model membranes. (A) Micelle, (B) bilayer

2.8 Neuropeptide receptors: Aim and scope of the investigation

Many neuropeptide systems consist of several closely related peptides that bind to more than one receptor subtype. Such a multireceptor/multiligand system is the Y-receptor family. They belong to the rhodopsin-like super family GPCRs, which are activated by the closely related peptide hormones neuropeptide Y, peptide YY and pancreatic polypeptide PP[93]. At least five Y-receptors have been cloned from mammals (Y1, Y2, Y4, Y5, and y6)[94]. The y6 receptor is designated in a lower case due to the fact that y6 is a truncated version of Y-receptors in mammals and no physiological function has been associated with it so far[95]. The major characteristic

features of these receptors are summarized in the Table 6. They are involved in control of a diverse set of behavioral processes including appetite, circadian rhythm, and anxiety[96, 97]; [98]. These receptors are expressed in the hypothalamus, in the brain stem and in peripheral tissues, such as blood vessels, lung, kidney, adrenal glands, stomach, colon, heart, pancreas and intestine[99-101]. An unusual feature of the Y-receptor family is their lack of sequence identity, particularly between Y1, Y2 and Y5 receptors, that are only 30% identical. A high homology of the Y1 to the Y4 (42%) and y6 receptors (51%) was found. The transmembrane regions share 40–43% identity. Subtypes of the Y1 receptor are known to exist in fish but not in mammals, while a further Y2-like receptor, named Y7, exists in amphibia, sharks and bony fish, but has apparently been lost from the mammalian lineage[94].

Receptor	**Y_1**	**Y_2**	**Y_4**	**Y_5**	**y_6**
Ligands	NPY, [P34]-NPY, PYY>>NPY/PYY fragments, PP	PYY, NPY, NPY (2–36)>>[P34]-NPY, PP	PP>>PYY>NPY ≥[P34]-NPY >>NPY/PYY fragments	NPY,PYY, [P34]-NPY, NPY(2–36), NPY(3–36) >PP	NPY, PYY>PP
Amino acids	384	381	375	455	371
Signal transduction	cAMP inhibition , Ca^{2+} mobilization				
Major Occurrence	Periphery, hypothalamus	Brain, hippocampus	Intestine, colon	Hypothalamus	Not in human
Related action	Vasoconstriction, anxiolysis, food intake?	Memory, epilepsy, secretion	Gastro-intestinal regulation	Food intake, Epilepsy	Not known

Table 6: Characterisation of Y-receptor subtypes. Modified from the reference [102]

All Y-receptor subtypes have two extracellular cysteines that probably may form a disulfide bond between extracellular loops I and II as confirmed by the X-ray structure of bovine rhodopsin and adrenergic receptor[5, 20]. Furthermore, all Y-receptors have one or several consensus sites for N-linked glycosylation (Asn– X–Ser/Thr) and at least one cysteine in the cytoplasmic tail that probably anchors the tail to the inside of the membrane by palmitoylation[103]. Studies on the structure-affinity and structure-activity relationship of each receptor type with variety of peptide analogs, as well as on site-directed mutagenesis and with anti-receptor antibodies, helped to

characterize each receptor subtype concerning its interaction with the ligand and its biological function[102]. The Y-receptors act via pertussis toxin-sensitive G- proteins, like members of the G_i and G_o family. Therefore, their activation leads to the inhibition of adenylyl cyclase and consequently, to the inhibition of cAMP accumulation in tissues and cells. In addition to this, inhibition and stimulation of K^+ and Ca^{2+} channels have been observed in neurons[104] and in the vasculature[105]. Apart from the regulation of food uptake, recently Y-receptors have been found linked to neoplasia. Y receptors were also found to be overexpressed in a variety of human cancers, mainly expressed in specific endocrine tumors and epithelial malignancies as well as in embryonal tumors[106-108]. NPY receptor expressing tumors are promising candidates for an *in vivo* NPY receptor targeting with radiolabeled and cytotoxic NPY analogs, analogous to somatostatin receptor targeting[109, 110].

Evidence suggests that NPY and its receptors are involved in the pathophysiology of several disorders, such as the control of food intake, metabolic disorders, anxiety, seizures, memory, circadian rhythm, drug addiction, pain, cardiovascular diseases, rhinitis and endothelial cell dysfunctions. A better knowledge of structure and mechanism of action of these receptors is clearly beneficial for pharmaceutical sciences. Even though in the recent years there has been spectacular progress in the structure determination of GPCRs by X-ray crystallography, it still remains a daunting task to obtain high-quality crystals of membrane proteins.

Nuclear magnetic resonance (NMR) becomes a valuable alternative technique. The less stringent requirements on the sample have allowed NMR to be widely applied to structural studies on MPs[111-117]. With the advance of NMR spectroscopic techniques such as TROSY[118], the generally considered molecular weight limitation in NMR studies has been dramatically lifted. It is established that assigning the NMR spectra of proteins as part of complexes in the 100 kDa range is now feasible[119-121]. Purifying milligram amounts of active GPCRs has long been and still is considered a highly challenging goal to achieve even for the best skilled research teams. Thus taking into account the enormous problems associated with the expression, purification, and reconstitution of intact GPCRs into membrane-mimetic environments suitable for biophysical studies, investigations on suitably chosen fragments of these proteins seem to be well justified. A crucial issue which remains is that, whether these truncated

constructs are able to successfully mimic structural features of the intact GPCRs. Extensive studies had been carried out to address this issue, which led to the conclusion that when a fragment of a membrane protein exhibits secondary structure, that structure is similar to the structure of the intact protein[82, 122]. A mentioned already, significant number of studies provided structural information on fragments corresponding to sequences of GPCR helices, loops and termini. E.g, the *Saccharomyces cerevisiae* pheromone receptor[76], the cannabinoid receptor[77], and neurokinin-1 receptor[78]. Furthermore, in our group we recently determined the conformation of a polypeptide corresponding to the 7^{th} TM helix of the yeast Ste2p receptor extended by 40 residues from the cytosolic tail[123] when integrated into DPC micelles and a construct comprising the first 2 TM domains in LPPG micelles[83].

The N-terminal domains from family 1 GPCRs have received little attention, most likely because of their short length, usually less than 70 amino acids. However, recent studies have suggested a pivotal role of N termini from GPCRs of this class in ligand recognition and binding[124-126]. Since these Y- receptors belong to the type 1b GPCRs, the ligand binding site involves the N-terminal part and the extracellular loops of the receptor. In this work we focussed on the structural studies of the N-terminus of the four Y-receptors (Y1, Y2, Y4 and Y5) of the family 1b GPCR that is targeted by members of the NPY family. Biosynthetic routes for recombinant production of the polypeptides in isotopically labeled form were established. The structure and the topology of these N-terminal domains in the presence of DPC or SDS micelles was elucidated by high-resolution NMR techniques. The N-terminal domains from all Y-receptors are fully unstructured in aqueous solution. On the contrary, in the presence of phospholipid micelles, they form helical segments with variable degree of stability.

Furthermore we developed an expression and purification strategy for the production of double transmembrane fragments of Y4 receptor in an isotopically enriched form in *E. coli*. The expression yields were optimized using different strains, different growth temperatures and various induction parameters. Even though the non-fusion (direct expression) method worked well for the N-TM1-TM2 construct (comprising the N-terminal domain, the first two transmembrane (TM) helices and the first extracellular loop followed by a $(His)_6$ tag) but for the other double-transmembrane constructs without fusion partner there was no expression at all, most

likely because their sequence was much shorter, possibly also because the very hydrophilic N-terminal part was not present in these constructs. Therefore we tried several fusion partners like the Trp leader, KSI fusion and even the N-terminus of the Y4 receptor to optimize the expression yields of the double transmembrane constructs. The fusion protein was cleaved using chemical cleavage (CNBr cleavage) method as well as enzymatic cleavage using the C3 protease enzyme. Extensive detergent screening was performed to yield the high quality of spectra required for our analysis. For the N-TM1-TM2 construct with ^{2}H, ^{13}C and ^{15}N labeled protein, the backbone was assigned almost completely in the presence of 1% DPC/6% LPPG mixed detergent micelles by my colleague Dr. Chao Zou, and now we are trying assign it in organic solvent mixture (50%TFE - water mixture) to facilitate assignments. The organic solvent mixtures seem appealing for membrane proteins since they give narrow line shapes and better resolution due to the smaller effective molecular weight of the protein in the absence of detergents. The mixed organic solvents systems are considered as a "quasi–native" membrane environment and have been validated for several membrane proteins by comparing its structure in the native conditions.

For example, the solution NMR structure of transmembrane H^{+} transporting subunit c of the F_1F_0 ATP synthase in a single-phase solution of chloroform–methanol–water (4:4:1) has also been shown to mimic that of the protein in the native complex[127]. But one has to be very careful in validating the relevance of the structure. However, it is not planned to compute the structure in the organic solvent mixture but to rather use the assignments made in this environment to assign the NOESY spectra recorded in lipids. These findings together will set the stage for the determination of the 3D structure of these large domains of GPCR in micelles using high-resolution NMR.

2.9 References

1. Foord, S.M., *Receptor classification: post genome.* Curr Opin Pharmacol, 2002. **2**(5): p. 561-6.
2. Takeda, S., et al., *Identification of G protein-coupled receptor genes from the human genome sequence.* FEBS Lett, 2002. **520**(1-3): p. 97-101.
3. Wilson, S. and D. Bergsma, *Orphan G-protein coupled receptors: novel drug targets for the pharmaceutical industry.* Drug Des Discov, 2000. **17**(2): p. 105-14.
4. Kristiansen, K., *Molecular mechanisms of ligand binding, signaling, and regulation within the superfamily of G-protein-coupled receptors: molecular modeling and mutagenesis approaches to receptor structure and function.* Pharmacol Ther, 2004. **103**(1): p. 21-80.
5. Cherezov, V., et al., *High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor.* Science, 2007. **318**(5854): p. 1258-65.
6. Rasmussen, S.G., et al., *Crystal structure of the human beta2 adrenergic G-protein-coupled receptor.* Nature, 2007. **450**(7168): p. 383-7.
7. Warne, T., et al., *Structure of a beta1-adrenergic G-protein-coupled receptor.* Nature, 2008. **454**(7203): p. 486-91.
8. Jaakola, V.P., et al., *The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist.* Science, 2008. **322**(5905): p. 1211-7.
9. P, E., *Address in pathology on chemotherapeutics: scientific principles.* Lancet, 1913. **2**: p. 445-451.
10. Langley, J.N., *On the contraction of muscle, chiefly in relation to the presence of ;receptive' substances: Part IV. The effect of curari and of some other substances on the nicotine response of the sartorius and gastrocnemius muscles of the frog.* J Physiol, 1909. **39**(4): p. 235-95.
11. Robinson GA, B.R.S.E., *Adenyl cyclase as an adrenergic receptor.* Ann NY Acad Sci 1967. **139**: p. 703.
12. Hardman, J.G., G.A. Robison, and E.W. Sutherland, *Cyclic nucleotides.* Annu Rev Physiol, 1971. **33**: p. 311-36.
13. Rodbell, M., et al., *The glucagon-sensitive adenyl cyclase system in plasma membranes of rat liver. V. An obligatory role of guanylnucleotides in glucagon action.* J Biol Chem, 1971. **246**(6): p. 1877-82.
14. Ross, E.M. and A.G. Gilman, *Resolution of some components of adenylate cyclase necessary for catalytic activity.* J Biol Chem, 1977. **252**(20): p. 6966-9.
15. Gilman, A.G., *G proteins: transducers of receptor-generated signals.* Annu Rev Biochem, 1987. **56**: p. 615-49.
16. Cassel, D. and Z. Selinger, *Catecholamine-stimulated GTPase activity in turkey erythrocyte membranes.* Biochim Biophys Acta, 1976. **452**(2): p. 538-51.
17. Northup, J.K., et al., *Purification of the regulatory component of adenylate cyclase.* Proc Natl Acad Sci U S A, 1980. **77**(11): p. 6516-20.
18. Codina, J., et al., *Pertussis toxin substrate, the putative Ni component of adenylyl cyclases, is an alpha beta heterodimer regulated by guanine nucleotide and magnesium.* Proc Natl Acad Sci U S A, 1983. **80**(14): p. 4276-80.

19. Bokoch, G.M., et al., *Purification and properties of the inhibitory guanine nucleotide-binding regulatory component of adenylate cyclase.* J Biol Chem, 1984. **259**(6): p. 3560-7.
20. Palczewski, K., et al., *Crystal structure of rhodopsin: A G protein-coupled receptor.* Science, 2000. **289**(5480): p. 739-45.
21. Rosenbaum, D.M., et al., *GPCR engineering yields high-resolution structural insights into beta2-adrenergic receptor function.* Science, 2007. **318**(5854): p. 1266-73.
22. Hanson, M.A., et al., *A specific cholesterol binding site is established by the 2.8 A structure of the human beta2-adrenergic receptor.* Structure, 2008. **16**(6): p. 897-905.
23. Park, J.H., et al., *Crystal structure of the ligand-free G-protein-coupled receptor opsin.* Nature, 2008. **454**(7201): p. 183-7.
24. Scheerer, P., et al., *Crystal structure of opsin in its G-protein-interacting conformation.* Nature, 2008. **455**(7212): p. 497-502.
25. Seifert, R. and K. Wenzel-Seifert, *Constitutive activity of G-protein-coupled receptors: cause of disease and common property of wild-type receptors.* Naunyn Schmiedebergs Arch Pharmacol, 2002. **366**(5): p. 381-416.
26. Bockaert, J. and J.P. Pin, *Molecular tinkering of G protein-coupled receptors: an evolutionary success.* Embo J, 1999. **18**(7): p. 1723-9.
27. Marinissen, M.J. and J.S. Gutkind, *G-protein-coupled receptors and signaling networks: emerging paradigms.* Trends Pharmacol Sci, 2001. **22**(7): p. 368-76.
28. Fredriksson, R., et al., *The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints.* Mol Pharmacol, 2003. **63**(6): p. 1256-72.
29. Vassilatis, D.K., et al., *The G protein-coupled receptor repertoires of human and mouse.* Proc Natl Acad Sci U S A, 2003. **100**(8): p. 4903-8.
30. Spiegel, A.M., *Defects in G protein-coupled signal transduction in human disease.* Annu Rev Physiol, 1996. **58**: p. 143-70.
31. Spiegel, A.M. and L.S. Weinstein, *Inherited diseases involving g proteins and g protein-coupled receptors.* Annu Rev Med, 2004. **55**: p. 27-39.
32. Milligan, G., *Strategies to identify ligands for orphan G-protein-coupled receptors.* Biochem Soc Trans, 2002. **30**(4): p. 789-93.
33. Horn, F., et al., *GPCRDB information system for G protein-coupled receptors.* Nucleic Acids Res, 2003. **31**(1): p. 294-7.
34. Harmar, A.J., *Family-B G-protein-coupled receptors.* Genome Biol, 2001. **2**(12): p. REVIEWS3013.
35. Pin, J.P., T. Galvez, and L. Prezeau, *Evolution, structure, and activation mechanism of family 3/C G-protein-coupled receptors.* Pharmacol Ther, 2003. **98**(3): p. 325-54.
36. Lagerstrom, M.C. and H.B. Schioth, *Structural diversity of G protein-coupled receptors and significance for drug discovery.* Nat Rev Drug Discov, 2008. **7**(4): p. 339-57.
37. Tate, C.G. and R. Grisshammer, *Heterologous expression of G-protein-coupled receptors.* Trends Biotechnol, 1996. **14**(11): p. 426-30.
38. Sarramegna, V., et al., *Heterologous expression of G-protein-coupled receptors: comparison of expression systems fron the standpoint of large-scale production and purification.* Cell Mol Life Sci, 2003. **60**(8): p. 1529-46.
39. LaVallie, E.R. and J.M. McCoy, *Gene fusion expression systems in Escherichia coli.* Curr Opin Biotechnol, 1995. **6**(5): p. 501-6.

40. Weiss, H.M. and R. Grisshammer, *Purification and characterization of the human adenosine A(2a) receptor functionally expressed in Escherichia coli.* Eur J Biochem, 2002. **269**(1): p. 82-92.
41. White, J.F., et al., *Automated large-scale purification of a G protein-coupled receptor for neurotensin.* FEBS Lett, 2004. **564**(3): p. 289-93.
42. Krepkiy, D., et al., *Bacterial expression of functional, biotinylated peripheral cannabinoid receptor CB2.* Protein Expr Purif, 2006. **49**(1): p. 60-70.
43. Baneres, J.L., et al., *Structure-based analysis of GPCR function: conformational adaptation of both agonist and receptor upon leukotriene B4 binding to recombinant BLT1.* J Mol Biol, 2003. **329**(4): p. 801-14.
44. Luca, S., et al., *The conformation of neurotensin bound to its G protein-coupled receptor.* Proc Natl Acad Sci U S A, 2003. **100**(19): p. 10706-11.
45. David, N.E., et al., *Expression and purification of the Saccharomyces cerevisiae alpha-factor receptor (Ste2p), a 7-transmembrane-segment G protein-coupled receptor.* J Biol Chem, 1997. **272**(24): p. 15553-61.
46. Andersen, B. and R.C. Stevens, *The human D1A dopamine receptor: heterologous expression in Saccharomyces cerevisiae and purification of the functional receptor.* Protein Expr Purif, 1998. **13**(1): p. 111-9.
47. Weiss, H.M., et al., *Comparative biochemical and pharmacological characterization of the mouse 5HT5A 5-hydroxytryptamine receptor and the human beta2-adrenergic receptor produced in the methylotrophic yeast Pichia pastoris.* Biochem J, 1998. **330 (Pt 3)**: p. 1137-47.
48. Mazina, K.E., C.D. Strader, and T.M. Fong, *Expression and solubilization of a recombinant human neurokinin-1 receptor in insect cells.* J Recept Res, 1994. **14**(1): p. 63-73.
49. Guan, X.M., T.S. Kobilka, and B.K. Kobilka, *Enhancement of membrane insertion and function in a type IIIb membrane protein following introduction of a cleavable signal peptide.* J Biol Chem, 1992. **267**(31): p. 21995-8.
50. Hermans, E., *Generation of model cell lines expressing recombinant G-protein-coupled receptors.* Methods Mol Biol, 2004. **259**: p. 137-53.
51. Ishihara, G., et al., *Expression of G protein coupled receptors in a cell-free translational system using detergents and thioredoxin-fusion vectors.* Protein Expr Purif, 2005. **41**(1): p. 27-37.
52. Lundstrom, K., *Structural genomics of GPCRs.* Trends Biotechnol, 2005. **23**(2): p. 103-8.
53. Lefkowitz, R.J., *G protein-coupled receptors. III. New roles for receptor kinases and beta-arrestins in receptor signaling and desensitization.* J Biol Chem, 1998. **273**(30): p. 18677-80.
54. Unger, V.M., et al., *Arrangement of rhodopsin transmembrane alpha-helices.* Nature, 1997. **389**(6647): p. 203-6.
55. Mizobe, T., et al., *Arrangement of transmembrane domains in adrenergic receptors. Similarity to bacteriorhodopsin.* J Biol Chem, 1996. **271**(5): p. 2387-9.
56. Ovchinnikov Yu, A., *Rhodopsin and bacteriorhodopsin: structure-function relationships.* FEBS Lett, 1982. **148**(2): p. 179-91.
57. Hargrave, P.A., et al., *The structure of bovine rhodopsin.* Biophys Struct Mech, 1983. **9**(4): p. 235-44.
58. Yoshizawa, T. and G. Wald, *Pre-lumirhodopsin and the bleaching of visual pigments.* Nature, 1963. **197**: p. 1279-86.

59. Teller, D.C., et al., *Advances in determination of a high-resolution three-dimensional structure of rhodopsin, a model of G-protein-coupled receptors (GPCRs).* Biochemistry, 2001. **40**(26): p. 7761-72.
60. Topiol, S. and M. Sabio, *X-ray structure breakthroughs in the GPCR transmembrane region.* Biochem Pharmacol, 2009. **78**(1): p. 11-20.
61. Stenkamp, R.E., D.C. Teller, and K. Palczewski, *Rhodopsin: a structural primer for G-protein coupled receptors.* Arch Pharm (Weinheim), 2005. **338**(5-6): p. 209-16.
62. Murakami, M. and T. Kouyama, *Crystal structure of squid rhodopsin.* Nature, 2008. **453**(7193): p. 363-7.
63. Milligan, G., P. Svoboda, and C.M. Brown, *Why are there so many adrenoceptor subtypes?* Biochem Pharmacol, 1994. **48**(6): p. 1059-71.
64. Chen-Izu, Y., et al., *G(i)-dependent localization of beta(2)-adrenergic receptor signaling to L-type Ca(2+) channels.* Biophys J, 2000. **79**(5): p. 2547-56.
65. Taylor, M.R., *Pharmacogenetics of the human beta-adrenergic receptors.* Pharmacogenomics J, 2007. **7**(1): p. 29-37.
66. Shi, L., et al., *Beta2 adrenergic receptor activation. Modulation of the proline kink in transmembrane 6 by a rotamer toggle switch.* J Biol Chem, 2002. **277**(43): p. 40989-96.
67. Rosenbaum, D.M., S.G. Rasmussen, and B.K. Kobilka, *The structure and function of G-protein-coupled receptors.* Nature, 2009. **459**(7245): p. 356-63.
68. Vogel, R., et al., *Functional role of the "ionic lock"--an interhelical hydrogen-bond network in family A heptahelical receptors.* J Mol Biol, 2008. **380**(4): p. 648-55.
69. Alkhalfioui, F., T. Magnin, and R. Wagner, *From purified GPCRs to drug discovery: the promise of protein-based methodologies.* Curr Opin Pharmacol, 2009. **9**(5): p. 629-35.
70. Meyer, B.H., et al., *FRET imaging reveals that functional neurokinin-1 receptors are monomeric and reside in membrane microdomains of live cells.* Proc Natl Acad Sci U S A, 2006. **103**(7): p. 2138-43.
71. Shcherbakova, O.G., et al., *Organization of beta-adrenoceptor signaling compartments by sympathetic innervation of cardiac myocytes.* J Cell Biol, 2007. **176**(4): p. 521-33.
72. Brum, P.C., et al., *Differential targeting and function of alpha2A and alpha2C adrenergic receptor subtypes in cultured sympathetic neurons.* Neuropharmacology, 2006. **51**(3): p. 397-413.
73. Pin, J.P., et al., *Allosteric functioning of dimeric class C G-protein-coupled receptors.* Febs J, 2005. **272**(12): p. 2947-55.
74. White, J.H., et al., *Heterodimerization is required for the formation of a functional GABA(B) receptor.* Nature, 1998. **396**(6712): p. 679-82.
75. Margeta-Mitrovic, M., Y.N. Jan, and L.Y. Jan, *A trafficking checkpoint controls GABA(B) receptor heterodimerization.* Neuron, 2000. **27**(1): p. 97-106.
76. Estephan, R., et al., *Biosynthesis and NMR analysis of a 73-residue domain of a Saccharomyces cerevisiae G protein-coupled receptor.* Biochemistry, 2005. **44**(35): p. 11795-810.
77. Ulfers, A.L., et al., *Structure of the third intracellular loop of the human cannabinoid 1 receptor.* Biochemistry, 2002. **41**(38): p. 11344-50.
78. Ulfers, A.L., A. Piserchio, and D.F. Mierke, *Extracellular domains of the neurokinin-1 receptor: structural characterization and interactions with substance P.* Biopolymers, 2002. **66**(5): p. 339-49.

79. Popot, J.L. and D.M. Engelman, *Helical membrane protein folding, stability, and evolution.* Annu Rev Biochem, 2000. **69**: p. 881-922.
80. Katragadda, M., J.L. Alderfer, and P.L. Yeagle, *Solution structure of the loops of bacteriorhodopsin closely resembles the crystal structure.* Biochim Biophys Acta, 2000. **1466**(1-2): p. 1-6.
81. Katragadda, M., J.L. Alderfer, and P.L. Yeagle, *Assembly of a polytopic membrane protein structure from the solution structures of overlapping peptide fragments of bacteriorhodopsin.* Biophys J, 2001. **81**(2): p. 1029-36.
82. Katragadda, M., et al., *Structures of the transmembrane helices of the G-protein coupled receptor, rhodopsin.* J Pept Res, 2001. **58**(1): p. 79-89.
83. Neumoin, A., et al., *Structure of a double transmembrane fragment of a G-protein-coupled receptor in micelles.* Biophys J, 2009. **96**(8): p. 3187-96.
84. Thevenin, D., et al., *Identifying interactions between transmembrane helices from the adenosine A2A receptor.* Biochemistry, 2005. **44**(49): p. 16239-45.
85. Thevenin, D., et al., *Oligomerization of the fifth transmembrane domain from the adenosine A2A receptor.* Protein Sci, 2005. **14**(8): p. 2177-86.
86. Dawson, J.P., J.S. Weinger, and D.M. Engelman, *Motifs of serine and threonine can drive association of transmembrane helices.* J Mol Biol, 2002. **316**(3): p. 799-805.
87. Yu, H., et al., *A general method for mapping tertiary contacts between amino acid residues in membrane-embedded proteins.* Biochemistry, 1995. **34**(46): p. 14963-9.
88. Sanders, C.R., 2nd and G.C. Landis, *Reconstitution of membrane proteins into lipid-rich bilayered mixed micelles for NMR studies.* Biochemistry, 1995. **34**(12): p. 4030-40.
89. Sanders, C.R. and R.S. Prosser, *Bicelles: a model membrane system for all seasons?* Structure, 1998. **6**(10): p. 1227-34.
90. Rastogi, V.K. and M.E. Girvin, *Structural changes linked to proton translocation by subunit c of the ATP synthase.* Nature, 1999. **402**(6759): p. 263-8.
91. Franzin, C.M., et al., *NMR of membrane proteins in micelles and bilayers: the FXYD family proteins.* Methods, 2007. **41**(4): p. 398-408.
92. Pervushin, K.V., et al., *Three-dimensional structure of (1-71)bacterioopsin solubilized in methanol/chloroform and SDS micelles determined by 15N-1H heteronuclear NMR spectroscopy.* Eur J Biochem, 1994. **219**(1-2): p. 571-83.
93. Michel, M.C., et al., *XVI. International Union of Pharmacology recommendations for the nomenclature of neuropeptide Y, peptide YY, and pancreatic polypeptide receptors.* Pharmacol Rev, 1998. **50**(1): p. 143-50.
94. Larhammar, D. and E. Salaneck, *Molecular evolution of NPY receptor subtypes.* Neuropeptides, 2004. **38**(4): p. 141-51.
95. Gregor, P., et al., *Molecular characterization of a second mouse pancreatic polypeptide receptor and its inactivated human homologue.* J Biol Chem, 1996. **271**(44): p. 27776-81.
96. Parker, E., M. Van Heek, and A. Stamford, *Neuropeptide Y receptors as targets for anti-obesity drug development: perspective and current status.* Eur J Pharmacol, 2002. **440**(2-3): p. 173-87.
97. Heilig, M., *The NPY system in stress, anxiety and depression.* Neuropeptides, 2004. **38**(4): p. 213-24.

98. Berglund, M.M., et al., *Studies of the human, rat, and guinea pig Y4 receptors using neuropeptide Y analogues and two distinct radioligands.* Peptides, 2001. **22**(3): p. 351-6.
99. Parker, R.M. and H. Herzog, *Regional distribution of Y-receptor subtype mRNAs in rat brain.* Eur J Neurosci, 1999. **11**(4): p. 1431-48.
100. Larsson, T.A., et al., *Characterization of NPY receptor subtypes Y2 and Y7 in rainbow trout Oncorhynchus mykiss.* Peptides, 2006. **27**(6): p. 1320-7.
101. Goumain, M., et al., *Identification and distribution of mRNA encoding the Y1, Y2, Y4, and Y5 receptors for peptides of the PP-fold family in the rat intestine and colon.* Biochem Biophys Res Commun, 1998. **247**(1): p. 52-6.
102. Cabrele, C. and A.G. Beck-Sickinger, *Molecular characterization of the ligand-receptor interaction of the neuropeptide Y family.* J Pept Sci, 2000. **6**(3): p. 97-122.
103. Holliday, N.D. and H.M. Cox, *Control of signalling efficacy by palmitoylation of the rat Y1 receptor.* Br J Pharmacol, 2003. **139**(3): p. 501-12.
104. Ewald, D.A., P.C. Sternweis, and R.J. Miller, *Guanine nucleotide-binding protein Go-induced coupling of neuropeptide Y receptors to Ca2+ channels in sensory neurons.* Proc Natl Acad Sci U S A, 1988. **85**(10): p. 3633-7.
105. Michel, M.C. and W. Rascher, *Neuropeptide Y: a possible role in hypertension?* J Hypertens, 1995. **13**(4): p. 385-95.
106. Kitlinska, J., et al., *Differential effects of neuropeptide Y on the growth and vascularization of neural crest-derived tumors.* Cancer Res, 2005. **65**(5): p. 1719-28.
107. Korner, M., B. Waser, and J.C. Reubi, *High expression of neuropeptide y receptors in tumors of the human adrenal gland and extra-adrenal paraganglia.* Clin Cancer Res, 2004. **10**(24): p. 8426-33.
108. Korner, M., B. Waser, and J.C. Reubi, *Neuropeptide Y receptor expression in human primary ovarian neoplasms.* Lab Invest, 2004. **84**(1): p. 71-80.
109. Reubi, J.C., et al., *High expression of peptide receptors as a novel target in gastrointestinal stromal tumours.* Eur J Nucl Med Mol Imaging, 2004. **31**(6): p. 803-10.
110. Korner, M. and J.C. Reubi, *NPY receptors in human cancer: A review of current knowledge.* Peptides, 2007. **28**(2): p. 419-25.
111. MacKenzie, K.R., J.H. Prestegard, and D.M. Engelman, *A transmembrane helix dimer: Structure and implications.* Science, 1997. **276**: p. 131.
112. Getmanova, E., et al., *NMR spectroscopy of phosphorylated wild-type rhodopsin: mobility of the phosphorylated C-terminus of rhodopsin in the dark and upon light activation.* Biochemistry, 2004. **43**(4): p. 1126-33.
113. Klein-Seetharaman, J., et al., *Differential dynamics in the G protein-coupled receptor rhodopsin revealed by solution NMR.* Proc Natl Acad Sci U S A, 2004. **101**(10): p. 3409-13.
114. Klein-Seetharaman, J., et al., *Solution NMR spectroscopy of [alpha -15N]lysine-labeled rhodopsin: The single peak observed in both conventional and TROSY-type HSQC spectra is ascribed to Lys-339 in the carboxyl-terminal peptide sequence.* Proc Natl Acad Sci U S A, 2002. **99**(6): p. 3452-7.
115. Schubert, M., et al., *Heteronuclear multidimensional NMR spectroscopy of solubilized membrane proteins: resonance assignment of native bacteriorhodopsin.* Chembiochem, 2002. **3**(10): p. 1019-23.
116. Oxenoid, K., et al., *NMR assignments for a helical 40 kDa membrane protein.* J Am Chem Soc, 2004. **126**(16): p. 5048-9.

117. Tian, C., et al., *Solution NMR spectroscopy of the human vasopressin V2 receptor, a G protein-coupled receptor.* J Am Chem Soc, 2005. **127**(22): p. 8010-1.
118. Pervushin, K., et al., *Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution.* Proc Natl Acad Sci U S A, 1997. **94**(23): p. 12366-71.
119. Fernandez, C. and G. Wider, *TROSY in NMR studies of the structure and function of large biological macromolecules.* Curr Opin Struct Biol, 2003. **13**(5): p. 570-80.
120. Trbovic, N., et al., *Efficient strategy for the rapid backbone assignment of membrane proteins.* J Am Chem Soc, 2005. **127**(39): p. 13504-5.
121. Tugarinov, V., et al., *Solution NMR-derived global fold of a monomeric 82-kDa enzyme.* Proc Natl Acad Sci U S A, 2005. **102**(3): p. 622-7.
122. Yeagle, P.L., G. Choi, and A.D. Albert, *Studies on the structure of the G-protein-coupled receptor rhodopsin including the putative G-protein binding site in unactivated and activated forms.* Biochemistry, 2001. **40**(39): p. 11932-7.
123. Neumoin, A., et al., *NMR studies in dodecylphosphocholine of a fragment containing the seventh transmembrane helix of a G-protein-coupled receptor from Saccharomyces cerevisiae.* Biophys J, 2007. **93**(2): p. 467-82.
124. DeMartino, J.A., et al., *The amino terminus of the human C5a receptor is required for high affinity C5a binding and for receptor activation by C5a but not C5a analogs.* J Biol Chem, 1994. **269**(20): p. 14446-50.
125. Dettin, M., et al., *CCR5 N-terminus peptides enhance X4 HIV-1 infection by CXCR4 up-regulation.* Biochem Biophys Res Commun, 2003. **307**(3): p. 640-6.
126. Ho, H.H., D. Du, and M.C. Gershengorn, *The N terminus of Kaposi's sarcoma-associated herpesvirus G protein-coupled receptor is necessary for high affinity chemokine binding but not for constitutive activity.* J Biol Chem, 1999. **274**(44): p. 31327-32.
127. Girvin, M.E., et al., *Solution structure of the transmembrane H+-transporting subunit c of the F1F0 ATP synthase.* Biochemistry, 1998. **37**(25): p. 8817-24.

Chapter 3

Expression of sCD14 (1-152) in *E.coli* and *Pichia pastoris*

3.0 Introduction

Protein–carbohydrate interactions are at the heart of many important biological processes including signaling, recognition and catalysis. The structural details of carbohydrate–protein complexes examined by NMR provide site-specific information on the architecture, binding selectivity and plasticity of the carbohydrate-binding sites of the proteins. So far, structural information at atomic resolution is only available on either the carbohydrate receptors, or on isolated (soluble) oligosaccharides and receptor–carbohydrate complexes, which are not membrane associated. Significant changes occur in the conformation especially in the membrane water interfacial region underlining the necessity for the study of protein carbohydrate interactions in this environment. This project aims at yielding structural data on the membrane-coupled carbohydrates and to gain insight into interactions with their ligands.

We have focused for our study on two systems:

1. Binding of sCD14(1-152) to lipopolysaccharides (LPS).
2. Binding of different types of polymyxins to lipopolysaccharides (LPS).

3.1 LPS

LPS is a major component of the outer membrane of Gram-negative bacteria, contributing greatly to the structural integrity of the bacteria. LPS also increases the negative charge of the cell membrane and helps stabilizing the overall membrane structure. It comprises three parts:1. The O-antigen (or O-polysaccharide) 2.The core oligosaccharide 3. LipidA. The lipidA portion of LPS inserts into the phospholipidic membrane and in many gram-negative bacteria consists of a di-glucosamine diphosphate with 5 to 7 fatty acid chains extending to one side of the disaccharide. The lipid A is appended to a region (the inner core) of 8-12 variable sugars (including the negatively charged 3-deoxy-D-

manno-oct-2-ulosonate (Kdo) units) and 3-8 phosphate residues. To the inner core is covalently associated the O-antigen, an oligosaccharide chain of variable length and chemical composition, depending on the exact type of LPS. Lipopolysaccharides are of crucial importance to gram-negative bacteria, and are therefore candidate targets for new antimicrobial agents.

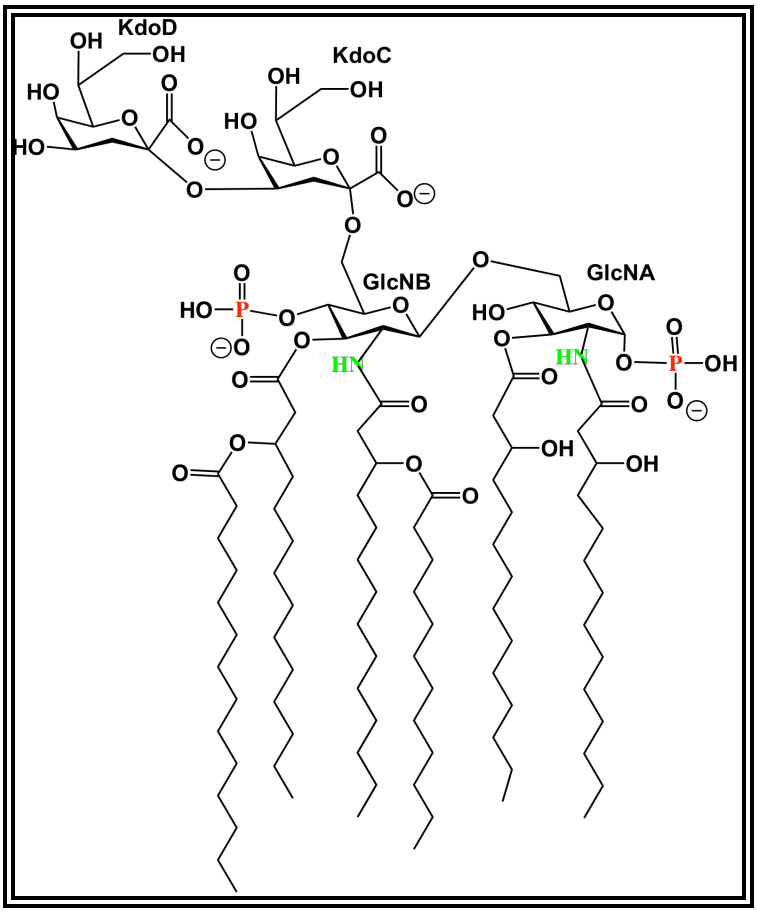


Figure 1: Chemical structure of E. coli Re LPS.
Kdo, 2-Keto-3-deoxyoctonate ; GlcN, N-acetylglucosamine.

If a wild strain of bacterium is irradiated with UV light or exposed to mutagenic compounds, it will mutate and LPS with shorter polysaccharide chains Ra, Rb, Rc, Rd, Re, etc may be formed where a, b, c, etc. designate 1st, 2nd, 3rd, etc. degree of the polysaccharide length of a given LPS. Ra and Re designate the mutants with the longest and shortest chain lengths, respectively. The most extreme mutants are the Re mutants

which produce an LPS which is made up of lipidA and 3-deoxy-D-manno-octulosonic acid (2-Keto-3-deoxyoctonate, KDO) as the sole constituent of the core[1](Figure1).

3.2 CD14

CD14 is a myelomonocytic differentiation antigen in mammals[2]. Human CD14 is a single-copy gene encoding two protein forms: a 50–55-kDa glycosylphosphatidyl-inositol (GPI)-anchored membrane protein (mCD14) and a monocyte or liver-derived soluble serum protein (sCD14) that lacks the anchor[3]. CD14 is the lipopolysaccharide (LPS)-LPS binding protein (LBP) receptor, forming a multi-protein complex containing at least CD14, MD-2 and TLR4[4]. The crucial role of CD14 in LPS signalling has been confirmed with knock-out mice; CD14-deficient mice are highly resistant to septic shock initiated by injection of either LPS or live bacteria[5]. Both mCD14 and sCD14 are critical for LPS-dependent signal transduction, and sCD14 confers LPS sensitivity to cells lacking mCD14. Mutant sCD14 containing only the N-terminal 152 amino acid residues were found to have activity equivalent to the full length sCD14[6].

The monomeric subunit of CD14 contains thirteen β strands, and 11 of them, from β-3 to β-13, overlap with conserved leucine-rich repeats (LRRs) (Figure 2). The concave surface of the horse-shoe-shaped structure consists of a large β-sheet of 11 parallel and two antiparallel beta strands. The convex surface of CD14 contains both helices and loops, in no regular pattern. A large hydrophobic pocket was found on the NH_2-terminal side. The pocket is the main component of the lipopolysaccharide-binding site. Dimerization in the crystal is mediated by residues in β 13 and in the loop between β 12 and β 13. Parallel β-sheets from the two monomers interact in an antiparallel fashion and form a large and continuous β-sheet encompassing the entire CD14 dimer[7]. CD14 is the primary receptor of the LPS signaling pathway, which leads to fatal septic shock syndrome. The structure provides evidence that different regions around the NH2-terminal pocket contribute to LPS binding and signaling. To this end, we wanted to study the interaction of sCD14(1-152) with LPS in the presence of DPC micelles. I will explain briefly about the attempts which were made to express sCD14(1-152) in a properly folded form (see Table-1).

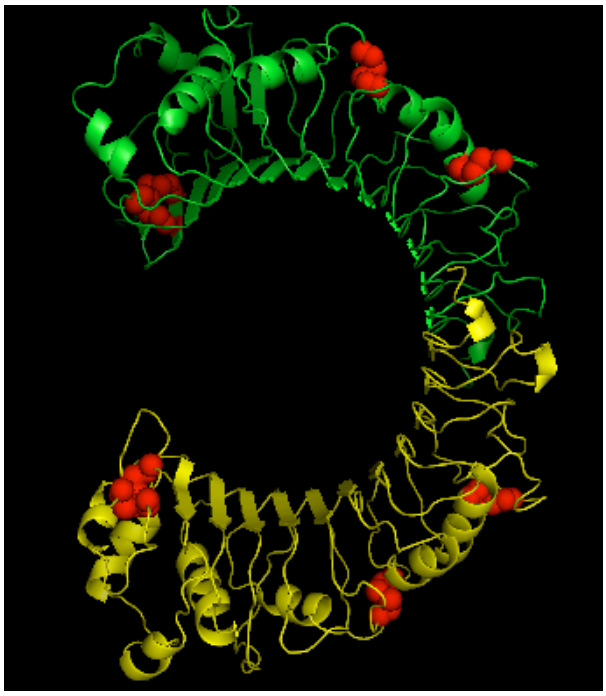


Figure 2: Structure of the CD14 dimer. Two monomers of CD14 in the crystal are colored in *green* and yellow. Disulfide bridges are shown as red spheres. mCD14 is GPI anchored membrane protein. Soluble sCD14 lacks the GPI anchor and mutant sCD14 containing N-terminal 152 amino acid residues sCD14 were found to have activity equivalent to the full length sCD14 and this construct was used for our studies.

3.3 Summary of expression of sCD14 (1-152)

3.3.1 Recombinant expression of sCD14(1-152) in *E.coli*

Escherichia coli is one of the most widely used host for protein expression because of its relative simplicity, its inexpensive and fast high-density cultivation, the well-known genetics and the large number of compatible tools available for biotechnology. Although often simple for soluble proteins, major obstacles are encountered in the expression of many heterologous proteins but a considerable amount of effort has been directed at improving the performance and versatility of this workhorse microorganism. We used C41(DE3) strain as the expression host for all the constructs as transformation with other strains like BL21(DE3) and BL21(DE3)pLysS was not successful. The proteins were purified from inclusion bodies under denaturing conditions.

The group of Jerala was able to express the recombinant CD14 and a CD14 fragment consisting of its first 152 residues in bacterial and the yeast *Pichia pastoris* expression system. Both bacterial and yeast recombinant CD14s were able to bind lipopolysaccharides with high affinity by dot blot assay and native PAGE gel shift assay. Initially, sCD14(1-152) was produced by Jerala for NMR studies[8, 9]. Unfortunately, although the protein was not precipitating NMR studies quickly revealed that the protein was heavily aggregated and not amenable to solution NMR studies. To reduce the extent of aggregation we measured ^{15}N-labelled sCD14(1-152) in presence of DPC micelles (Figure 4). Unfortunately, only a fraction of the expected signals with largely varying line-widths are observed, and the signal dispersion was limited.

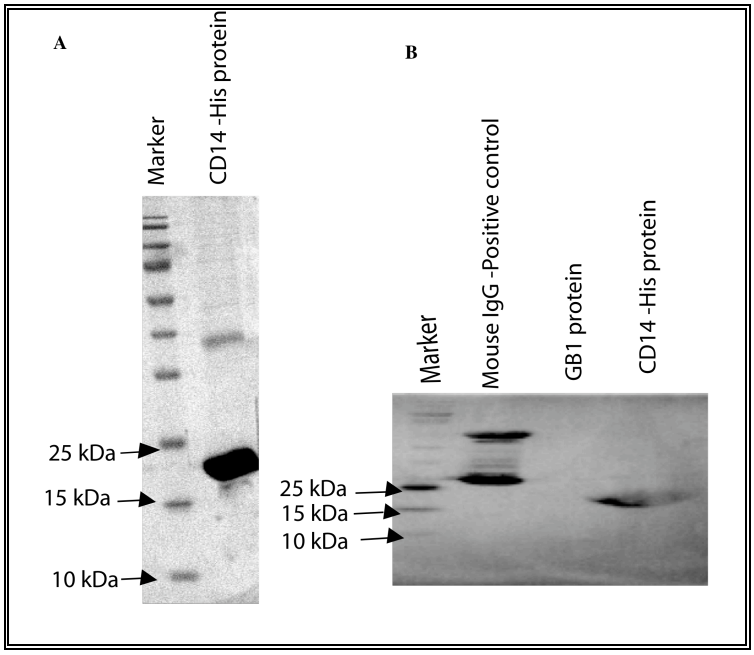


Figure 3 A: Pure CD14-His protein by SDS- PAGE analysis.

3 B: Western blotting of CD14-His protein with anti-human CD14 raised in mouse.

We suspected that our expression product did not display correctly formed disulfide bonds and therefore decided to use Origami host strains. Those are K-12 derivatives that have mutations in both the thioredoxin reductase (trxB) and glutathione reductase (gor) genes, which greatly enhances disulfide bond formation in the cytoplasm. We tried expression of sCD14(1-152) in OrigamiDE3 competent cells, which provides an

oxidative environment for the bacterial cytoplasm to allow formation of disulphide bonds, but the growth was very poor in M9 medium.

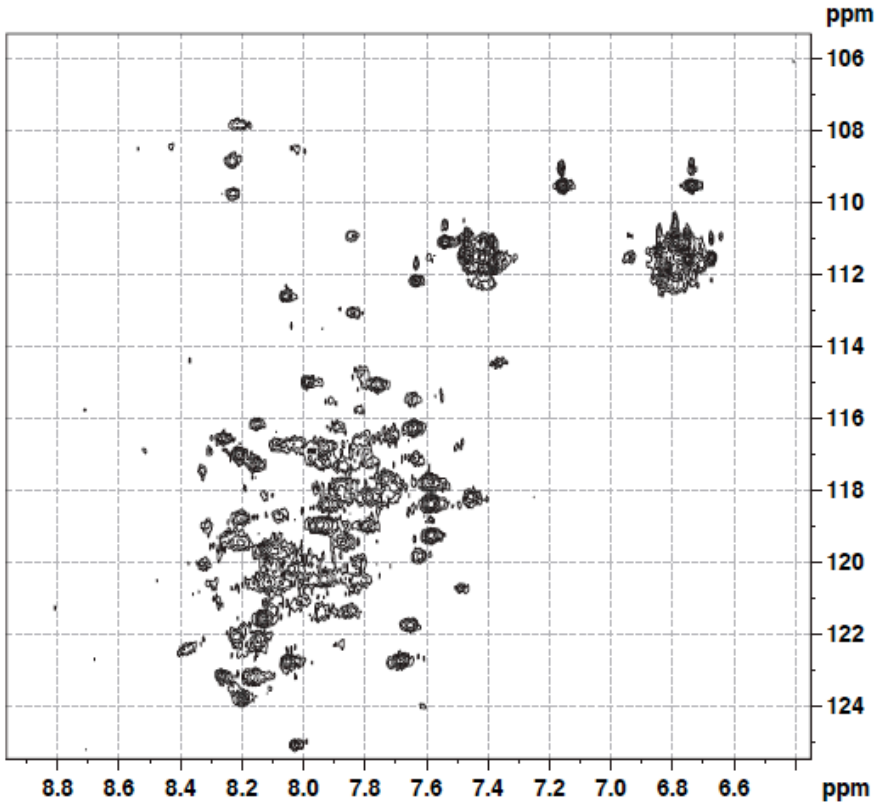


Figure 4: [^{15}N,^{1}H]-HSQC spectrum of 2 mM CD14-His protein in 300 mM d_{38}-DPC, 20 mM PO_4 Buffer pH 6.8, T=310K.

We therefore decided to attach a solubility enhancing tag to the protein. Wagner had been able to demonstrate that such a tag both increases levels of expression but also stabilizes the protein in monomeric state at much higher concentrations[10]. Expression vectors have been designed which encode the immunoglobulin–binding domain of streptococcal protein G (GB1; 56 residues) linked to the N-terminus of CD14 protein. This fusion strategy takes the advantage of small size, stable fold and high bacterial expression capability of the GB1 domain to allow direct NMR spectroscopic analysis without the need of the removal of the fusion protein. Moreover, the resonance positions of GB1 are known and assigned, and can thereby be distinguished from resonances of CD14. This construct (pET 28a His–GB1-$(SG)_2$CD14) again yielded aggregated protein

and hence was not pursued any further. The construction of functional fusion protein often requires a linker sequence that adopts an extended conformation to allow for maximal flexibility**.** In this construct (pET 28a-HisG$(SG)_6$CD14) the length of the serine glycine linker in between GB1 domain and CD14 was increased relative to the pET 28a-HisG$(SG)_2$CD14 construct. We increased the length of serine glycine linker from two to six by adding four additional aminoacids (two serine and two glycine residues were added). The protein expression is good (Figure 5A and B) but it runs in the SDS gel at a higher molecular weight. The mass is found to be correct by MS analysis. In this construct also the *in vitro* refolding was not fruitful.

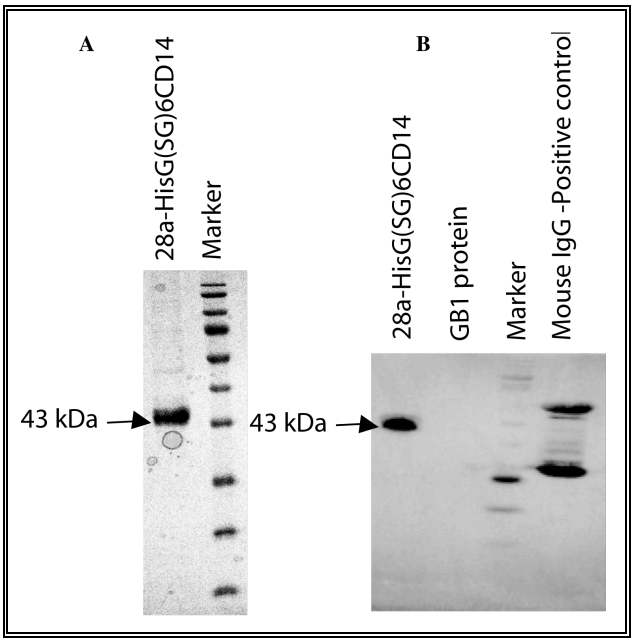


Figure 5 A: Pure 28a-His$(GS)_6$CD14 protein by SDS- PAGE analysis.
B: Western blotting of CD14-His protein with anti-human CD14 raised in mouse

The N-terminal 152 fragment of soluble CD14 is a truncated protein. On both ends of the protein it may need a soluble fusion tag to mask the exposed hydrophobic residues at the C-terminus of the protein in order to form a properly folded protein. We decided to add a C- terminal His tag to CD14. It yielded protein in the soluble fraction, but the [^{15}N,^{1}H]-HSQC revealed that only the GB1 domain was well-folded in the presence of 300 mM DPC (Figure 7).

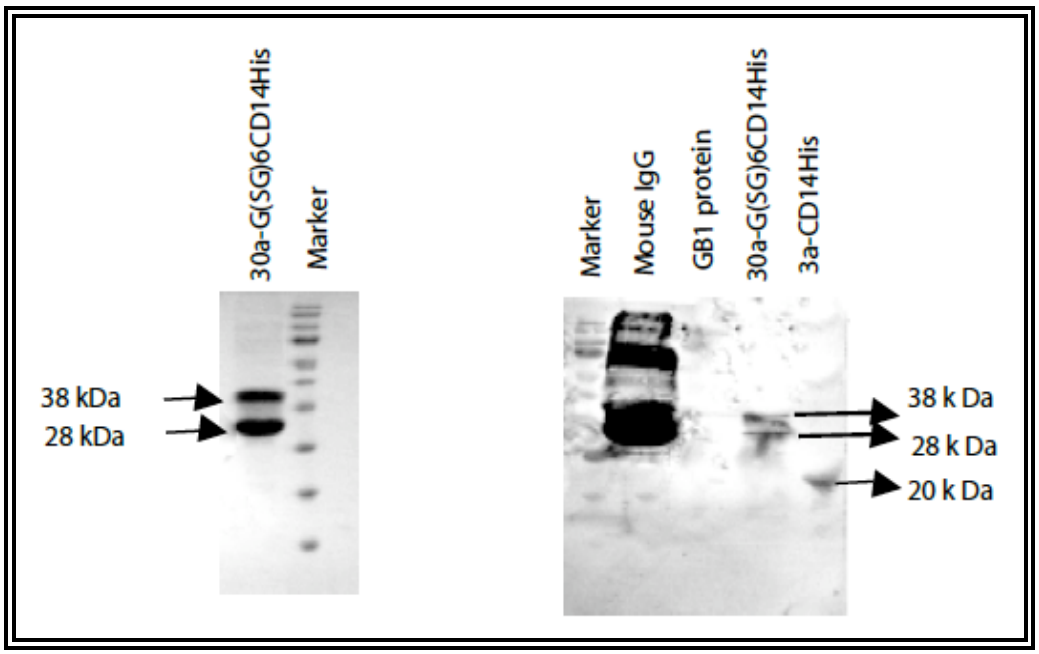


Figure 6 A: Pure 30a-G(SG)6CD14 His protein by SDS- PAGE analysis.

B: Western blotting 30a-G(SG)6CD14 His protein with anti-human CD14 raised in mouse.

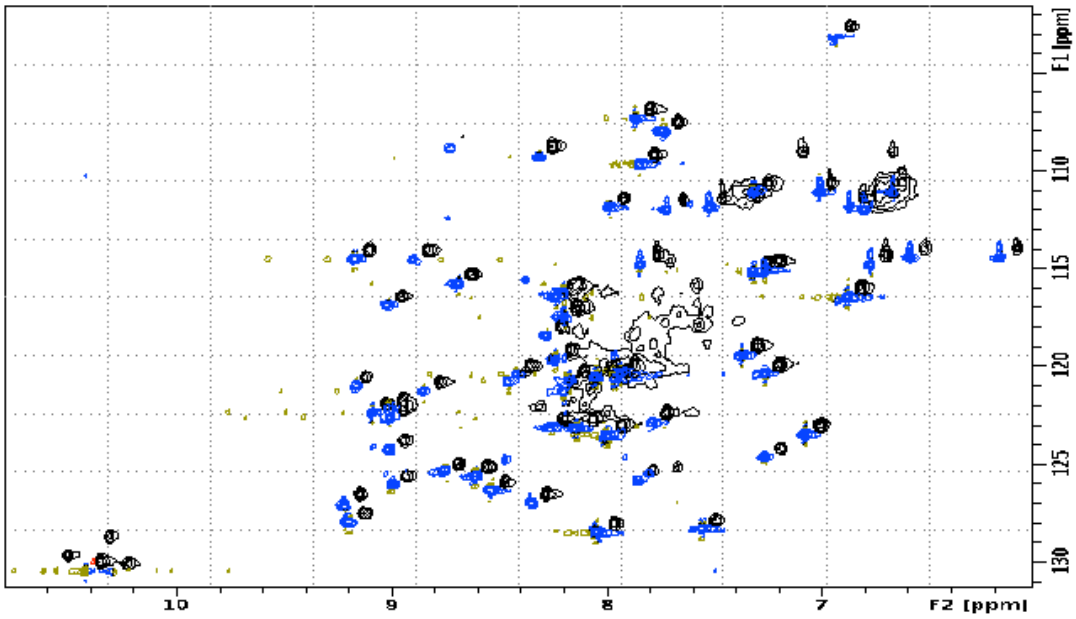


Figure 7: [^{15}N,^{1}H]-HSQC spectrum of 2 mM 30aG(SG)6CD14His protein. (black coloured peaks) overlayed with 0.3 mM GB1 protein(blue peaks) in 300 mM d_{38}-DPC, 20 mM PO_4 Buffer pH 6.8, T=310K.

Sometimes, folded states of proteins can be stabilized upon addition of their native ligand, and hence we added LPS, but no improvement was seen (Figure 8). A gel filtration chromatography preformed to check the aggregation state of the protein revealed that the protein eluted at a volume corresponding to the size of 110 kDa instead of 25 kDa while control proteins BSA and Carbonic anhydrase eluted at the expected volume. This indicates even though the protein was in the soluble fraction it was still in a highly aggregated state.

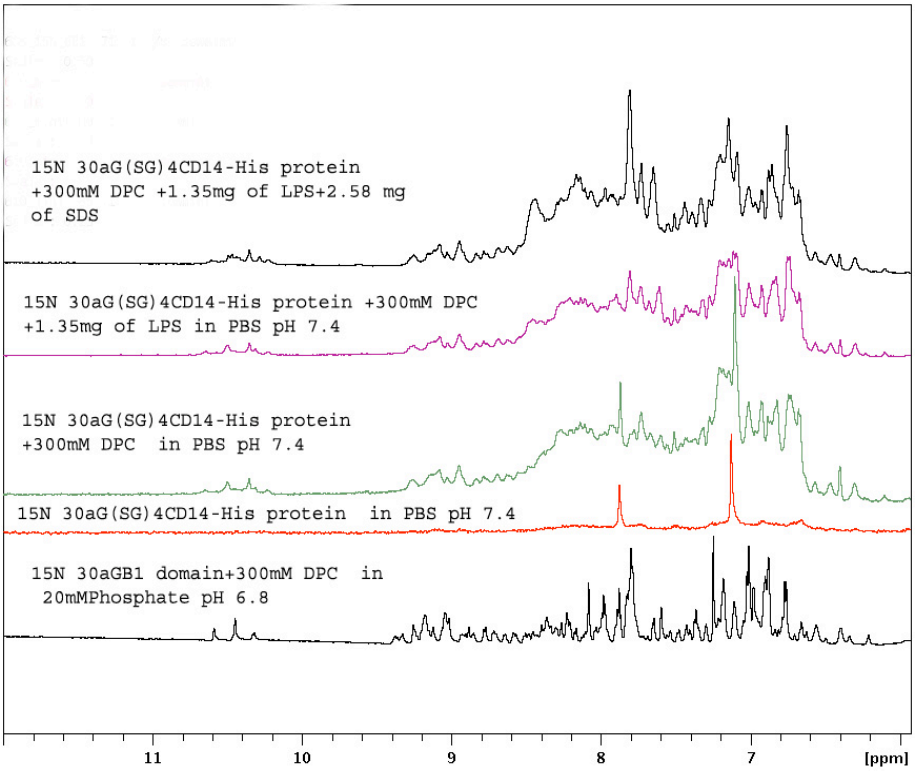


Figure 8: 1D ^{1}H NMR of ^{15}N 30a-G(SG)6CD14-His protein and ^{15}N 30aGB1 domain .

Expression of CD14 from *E.coli* cytoplasmic inclusion bodies lead to the production of insoluble and nonfunctional protein and *in vitro* refolding was not fruitful in our case. We therefore decided to direct these polypeptides to the bacterial periplasmic space, with the expectation to find better conditions for proper folding, especially for the formation of stabilizing disulphide bond. The periplasm is a compartment where oxidation of thiols can occur due to the activity of the disulphide oxidoreductase (Dsb) system, and there are fewer proteases in the periplasm compared to the cytoplasm and many have specific substrates. Finally, because the periplasm contains fewer protein,

easier purification of the target protein is facilitated. The protein (PelB-GB1-$(SG)_6$-CD14) was expressed in C41(DE3) cells at low temperature (22°C) to favor periplasmic expression and then purified by osmotic shock protocol, whereas PelB denotes the signaling sequence to direct the fusion for periplasmic expression. Unfortunately, no CD14 protein was contained in the soluble fraction, and the mass of the protein was also found to be wrong. Some contaminating periplasmic protein was co-purified, but did not display bands in the western blot using antiCD14 antibody for the soluble fraction. The western blot displayed a positive reaction in the cell pellet only. The reasons for this failure may be as follows

1. The osmotic shock, that releases proteins from the periplasmic space, might not have worked properly
2. There could be a problem in translocation of the fusion to the periplasmic space.
3. Since CD14 is a receptor for bacterial LPS there could be a problem of binding of CD14 to LPS from the cell membrane once it was released from the periplasmic space after the osmotic shock. In all the other constructs since it was purified under denaturing conditions, and the problem of binding to LPS might not be present under those circumstances.
4. No purification under denaturing condition was attempted since there was no affinity tag for further purification. Furthermore, no attempts to solubilize the protein from the membrane fraction using detergents were made because once the CD14 binds to LPS it would be very difficult to obtain the pure CD14 without the LPS contamination. We introduced a His tag at the C-terminus to the above pelB fusion in order to allow purification under denaturing condition. Unfortunately, this protein also could not be refolded properly.

We tried *in vitro* refolding for all these constructs with different buffers taking into consideration the following aspects: 1. Generation of the native structure requires both folding and concomitant formation of the correct disulfide bonds. 2. For the formation of proper disulphide bonds we have to provide a redox system in our refolding buffer to allow reshuffling of the disulphide bond. *In vitro* oxidation is a very slow process and sufficient time should be given for the formation of appropriate disulphide

bond. 3. Aggregation that is, formation of inactive, insoluble protein, is a major problem encountered upon reconstitution and to prevent this to occur low molecular weight additives like arginine, glutamine etc are included in the refolding buffer.

The following table -1 summarizes my attempts to express different constructs of CD14 in *E. coli*

Construct name	**Protein Expression details**
pET3a **I CD14-His** (Cloned between NdeI and BamH1)	**Purification** :by Ni.affinity chromatography under denaturing condition **Size :** 17.78 kDa **Yield:** 40 mg /litre **Solubility:** not Soluble in PO_4 buffer; soluble in DPC **NMR:** the construct looks like an aggregated sample **W.Blot:** positive reaction with antiCD14 antibody. **Refolding :** not successful **Other Strains:** Origami 2(DE3) – To promote the formation of appropriate disulphide bond formation we used this competent cells for expression. Transformation worked with commercial COM cells. Growth is fine in LB but no growth was observed in M9 medium **MS results:** Mass expected-17783.00 Da ; Mass obtained -17650.549 Da . Difference of 132.451 which corresponds to the mass of N-terminal Methionine (131 Da)
pET30a **II GB1-His** (GB1- Cloned between NdeI and BamH1	**Purification:** by Ni-affinity chromatography under native condition. **Size:** 7.18 kDa **Yield:** 35 mg /liter **Solubility:** Soluble in PO_4 buffer and DPC **NMR:** Obtained good spectrum. Looks like properly folded protein **W.Blot :** no reaction with antiCD14 antibodies **MS results:** Mass expected-1455.56 Da(5 times charged) ; Mass obtained -1455.1 Da . Difference - 0.4 Da
pET28a **III His G(SG)2CD14** GB1- Cloned between NdeI and BamH1 CD14 – Cloned between BamHI and EcoRI	I have mainly used this construct for further cloning **Purifcation** :by Ni-chromatography under denaturing condition **Result :** protein aggregated upon purification and no further refolding attempts were made
pET28a **IV His G$(SG)_6$CD14** GB1- Cloned between NdeI and BamH1 Six residues of Serine Glycine linker CD14 – Cloned between BamHI and EcoRI	**Purifcation:** by Ni- affinity chromatography under denaturing condition. **Size:** 25.76 kDa **Yield:** 30 mg /liter (based on the weight of precipitate) **Solubility:** not soluble in PO_4 buffer and DPC **NMR:** looks like aggregated sample **W.Blot:** positive reaction with antiCD14 antibody. **Refolding :** tried refolding similar to the pET3aCD14 construct – not successful **MS results:** Mass expected-25757.00 Da ; Mass obtained -25626.5 Da . Difference of 131.51 which corresponds to the mass of N-terminal Methionine (131 Da)
pET30a **V G$(SG)_6$CD14His** GB1- Cloned between NdeI and BamH1. Six residues of Serine Glycine linker CD14 – Cloned between BamHI and EcoRI	**Purification:** by Ni. affinity chromatography under denaturing condition. Size:24.42 kDa **Yield:** 30 mg /litre **Solubility**.soluble in PO_4 buffer and DPC **NMR:** in aqueous buffer it looks aggregated. When dissolved in DPC, a part of the protein resulted in a spectrum which corresponds to folded protein. But when comparing it to the spectrum of the GB1 domain there were no extra peaks corresponding to CD14. **W.Blot**: positive reaction with antiCD14 antibody. **Refolding:** not successful. **Gel filtration:** eluted at a volume corresponding to an aggregated sample

	MS results: Mass expected-24416.5 Da; Mass obtained -24403.3 Da . Difference - 13.2 Da .
pET26b **VII PelB-G(SG)$_6$CD14** GB1- Cloned between NCoI and BamH1 Six residues of Serine Glycine linker CD14– Cloned between BamHI andEcoRI. No His tag Pelb sequence cleaved in *E.coli* cell, when it is translocated to the cytoplasm	**Purification:** applied osmotic shock and separated the periplasmic fraction and then dialyzed the sample. No further purification done because that fraction itself did not contain any impurities. **Size**: 26.05 kDa **Solubility:** soluble in PO_4 buffer. **NMR:** no NMR data were recorded for this sample, because of the negative reaction in the Western blot and the mass also was found to be incorrect. **W.Blot:1)** negative for the soluble periplasmic extract. **2**) Positive reaction with antiCD14 antibody for the cell pellet after osmotic shock. **Conclusion:** the protein obtained in the soluble fraction was some periplasmic contaminating protein. **MS results:** Mass expected-26054.7 Da ; Mass obtained – 28161.00Da . Difference – 2107.00 Da. (some contaminating periplasmic protein) .
pET26b **VI Pelb-G(SG)$_6$CD14His** Pelb GB1(SG)$_6$CD14His GB1- Cloned between NCoI and BamH1 ;CD14 – Cloned between BamHI andEcoRI;C-Ter His Tag Pelb sequence cleaved in *E.coli* cell, when it is translocated to the cytoplasm	**Purification:** by Ni. affinity chromatography under denaturing condition. **Size**:27.6 kDa **Result:** Only precipitated protein from protein purification under denaturating conditions. **W.Blot:** Protein precipitate gave an positive reaction with antiCD14 antibody

Table 1: Summary of the different CD14 constructs in *E.coli* expression system.

Unfortunately, no constructs of well-folded protein in non-aggregated form could be produced, although some of them were expressed at reasonable yield. However, the NMR experiments quickly revealed that the resulting fusions were either unfolded or highly aggregated and hence not amenable to detailed studies by NMR. For example the [^{15}N,^{1}H]-HSQC of the CD14 construct without any fusion partner except C-terminal His tag (Figure 4) and an CD14 construct with N-terminal GB1 fusion and C-terminal His tag (Figure 7) overlayed with the properly folded GB1 domain protein spectra is shown. The poor spectral dispersion seen in the spectrum in Figure 4 reveals that it is most probably in the denatured state. In case of the spectrum depicted in Figure 7 the [^{15}N,^{1}H]-HSQC reveals that only the GB1 domain was well-folded and that CD14 still remains in the aggregated state in the presence of 300mM DPC. We speculated that the failure is largely due to the missing folding machinery of *E. coli.*

3.3.2 Recombinant protein expression in *Pichia Pastoris*

Although prokaryotic recombinant protein expression systems are often used in structural studies, mammalian proteins often accumulate in insoluble form. *In vitro* refolding of the protein is also not successful always. We decided to switch our expression host to *Pichia pastoris*. The Pichia expression system produces high yields of recombinant protein in a well-characterized, easy to manage eukaryotic host. The high yields with the Pichia expression system are achieved by the key factors like 1) high biomass. The cell density of *Pichia pastoris* can be 10 times greater than *Sacchoromyces cerevesia*. 2) Strong promoters –both inducible and constitutive promoters are available to drive high level expression. 3) Efficient secretion - recombinant proteins fused to the alpha factor or a native secretion signal will be directed into the medium, simplifying purification. Together these features give high levels of expression in a cost–effective microbial eukaryotic host.

3.3.2.1 The pichia expression vector

pPICZαC is a 3.6 kb vector used to express and secrete recombinant proteins in *Pichia pastoris*. Recombinant proteins are expressed as fusions to a N-terminal peptide encoding the *Saccharomyces cerevisiae* α -factor secretion signal. The Pichia expression vectors take advantage of the powerful AOX1 (Alcohol oxidase) promoter and uses methanol to induce high-level expression of the gene of interest. The constructs integrate into the *Pichia pastoris* genome, creating a stable host that generates extremely high protein expression levels and can be used in any Pichia strain including X-33, SMD1168H, and KM71H. The CD14 gene was cloned between Xho1 site and Xba 1 site of pPICZαC vector to include the alpha factor in our construct. But the CD14 gene has an internal Xho1 site. Site-directed mutagenesis was performed to mutate this site before cloning into pPICZαC vector. The alpha factor was responsible for targeting the protein into the medium and this was necessary for disulfide bond formation.

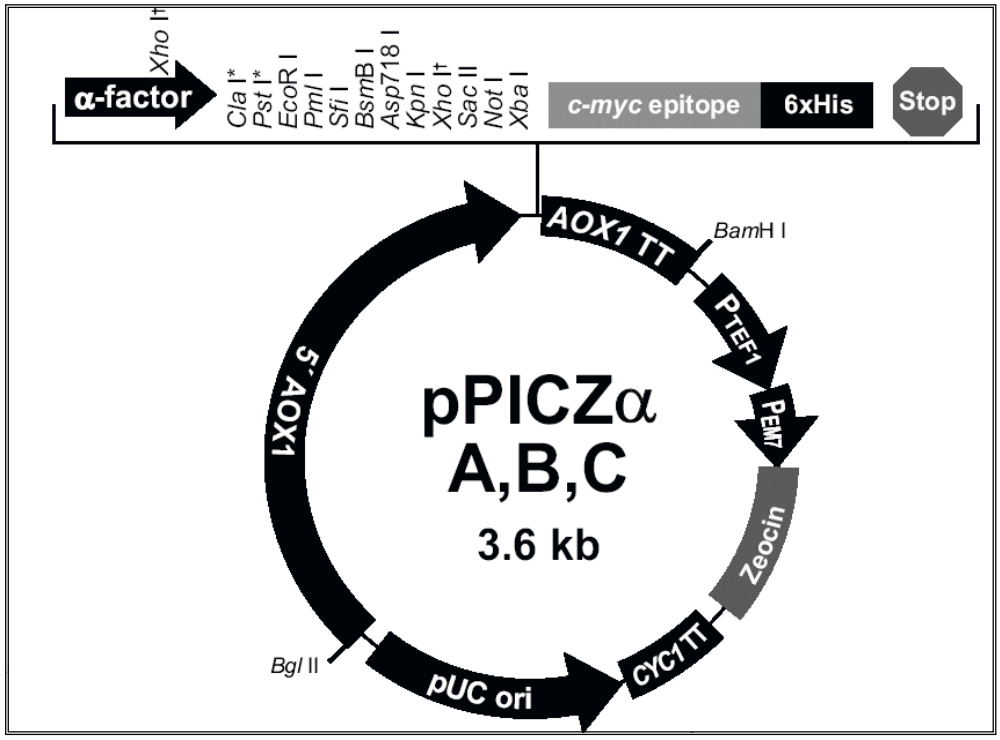


Figure 9: pPICZα plasmid map

3.3.2.2 How it works

The host for high-level recombinant protein expression in the Pichia expression system is the methylotrophic yeast *Pichia pastoris*. In the absence of glucose, *Pichia pastoris* uses methanol as the sole carbon source. The first step in the methanol metabolism is oxidation of methanol to formaldehyde; this reaction generates also hydrogen peroxide. To avoid hydrogen peroxide toxicity, methanol metabolism takes place inside the peroxisome, which sequesters the toxic by-products. Alcohol oxidase has a poor affinity for O_2 and Pichia compensates by producing the enzyme in large quantities. The alcohol oxidase (AOX1) promoter controls expression of alcohol oxidase, and is used to drive the heterologous protein expression in Pichia. Typically, 30% of the total soluble protein in methanol-induced cells is alcohol oxidase.

3.3.2.3 Experimental overview

The following table describes the basic steps needed to clone and express the gene of interest in pPICZαC vetor

Step	Action
1	Propagate pPICZαC by transformation into a recA, endA1 E. coli strain such as TOP10, DH5α, or JM109.
2	Develop a cloning strategy and ligate the gene into one of the pPICZ αC vectors in frame with the α-factor secretion signal and the C-terminal tag.
3	Transform into *E. coli* and select transformants on low-salt LB plates containing 25 μg/ml Zeocin.
4	Analyze transformants by restriction mapping or sequencing to confirm in-frame fusion of your gene with the α-factor secretion signal and also with C-terminal tag if desired.
5	Purify and linearize the recombinant plasmid for transformation into *Pichia pastoris*.
6	Transform in to the required Pichia strain and plate onto YPDS plates containing the appropriate concentration of Zeocin.
7	Select for Zeocin-resistant transformants.
8	Optimize expression of your gene.
9	Purification of the fusion protein by affinity purification.

Table-2: Overview of pichia expression system

3.3.2.4 Expression of CD14 in *Pichia pastoris*

After obtaining the clone, the pPICZαC+CD14 vector is completely linearized with *Pme* I restriction digestion enzyme which is present in the 5′ *AOX1* region of the vector. Vector linearized within the 5′ *AOX1* region will integrate by gene insertion into the *P.pastoris* host 5′ A*OX1* region. The linearized vector is subjected to ethanol precipitation and dried which was then transformed into the following three major strains of *P. pastoris* such as protease-deficient SMD1163 (*his*4, *pep*4, *pbr*), GS115 (his4+) and X-33 (wild-type), by electroporation method using freshly prepared competent cells. Incubate the plates for 2 to 3 days at 30°C until colonies form. Transformation resulted in around ten colonies with the GS115 strain and many colonies with the protease deficient strain SMD1163 and no colonies with X-33 strain.

Around 10-20 colonies from both the strains were streaked onto fresh YPDS plates (Yeast extract peptone dextrose medium+sorbitol) containing two different

concentration of zeocin (25ug/ml and 50ug/ml) to select more effectively that they were pure *Pichia* strains with the gene of the interest. Around 10 colonies for each strain was inoculated into 25ml BMGH (buffered complex glycerol histidine medium) for around 18 – 20 hrs at 30ºC (temperature and aeration are critical factors for growth of *pichia* strains). The cells were pelleted down and the pellet was resuspended into 100 -200 ml of BMMH (buffered minimal methanol histidine medium) medium (without dextrose) to give an OD_{600} 1. Then the culture was induced with 0.5% methanol every 24 hrs and allowed to grow at 30ºC for 72 hrs. 1ml of the expression culture was transferred to analyse expression level every 24 hrs. At the end of 72hrs, cells were then centrifuged, the supernatant was separated. Proteins secreted into the medium should be usually >50% homogenous and would require some additional purification. Since the expression level was very low, they are concentrated by TCA precipitation/ammonium sulphate precipitation. In most of the clones, no precipitate was obtained with the culture supernatant. In few, small amount of precipitate was obtained and the protein expression was checked by western blot (Figure 10) since I could not detect any expression by SDS PAGE electrophoresis.

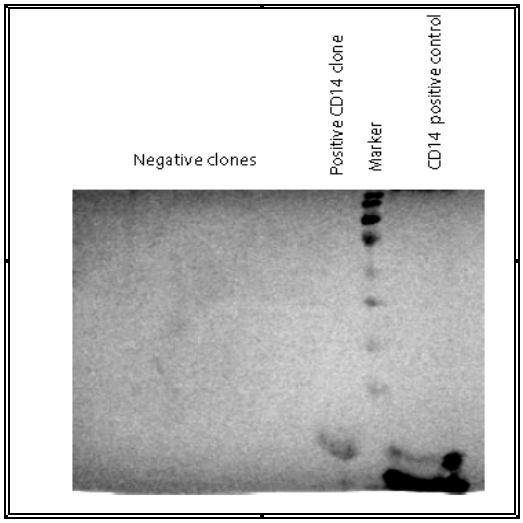


Figure 10: Western blotting to check the clones for CD14-His protein expression in *Pichia* strains with anti-human CD14 raised in mouse

Even after screening hundreds of colonies no clones with optimal expression of CD14 could be detected. (In very few clones expression could be detected by western

blot). This low expression level will not be of any help for our structural studies by NMR. The major reasons for the no/low expression could be that the protein is not secreted into the medium. But I have checked few of the cell pellets also. But I didn't see the expression of the protein. May be the protein is hyperglycosulated so that it cannot pass through the secretion pathway. Other potential reason could be, since the protein is excreted into medium there may be proteolysis of the protein or there is some problem in incorporation of our target gene into *pichia* genome or premature transcriptional termination of the target gene. Then we decided to move on to the next system, which was mainly focused to obtain experimental data on the LPS-Polymyxin complex, thereby allowing a detailed understanding of the interacting moieties.

3.3.2.5 Outlook

Since this construct sCD14 (1-152) does not give a properly folded protein, it is good to try the expression of constructs which are longer or shorter than this construct. It is worth trying making a series of peptides and screen them to asses their binding affinity for *E.coli* Re-LPS. By using a very soluble tag on both ends which then can be cleaved by using C3 protease or TEV protease enzymes could also help for the production of properly folded protein. It is good to try expression of other key players of LPS signaling pathway like BPI (bacterial permeability increasing protein) to study their interactions with Re-LPS. It was proved that the amino-terminal region of BPI,($BPI_{1–197}$, rBPI23) retains full LPS-binding activity[11].

3.5 References

1. Raetz, C.R., *Biochemistry of endotoxins*. Annu Rev Biochem, 1990. **59**: p. 129-70.
2. Schutt, C., *Cd14*. Int J Biochem Cell Biol, 1999. **31**(5): p. 545-9.
3. Haziot, A., et al., *The monocyte differentiation antigen, CD14, is anchored to the cell membrane by a phosphatidylinositol linkage*. J Immunol, 1988. **141**(2): p. 547-52.
4. Miyake, K., *Innate recognition of lipopolysaccharide by CD14 and toll-like receptor 4-MD-2: unique roles for MD-2*. Int Immunopharmacol, 2003. **3**(1): p. 119-28.
5. Haziot, A., et al., *CD14-deficient mice are exquisitely insensitive to the effects of LPS*. Prog Clin Biol Res, 1995. **392**: p. 349-51.

6. Juan, T.S., et al., *Soluble CD14 truncated at amino acid 152 binds lipopolysaccharide (LPS) and enables cellular response to LPS*. J Biol Chem, 1995. **270**(3): p. 1382-7.
7. Kim, J.I., et al., *Crystal structure of CD14 and its implications for lipopolysaccharide signaling*. J Biol Chem, 2005. **280**(12): p. 11347-51.
8. Majerle, A., J. Kidric, and R. Jerala, *Expression and refolding of functional fragments of the human lipopolysaccharide receptor CD14 in Escherichia coli and Pichia pastoris*. Protein Expr Purif, 1999. **17**(1): p. 96-104.
9. Majerle, A., J. Kidric, and R. Jerala, *Bacterial expression and refolding of different fragments of human CD14*. Pflugers Arch, 2000. **439**(3 Suppl): p. R109-10.
10. Zhou, P., A.A. Lugovskoy, and G. Wagner, *A solubility-enhancement tag (SET) for NMR studies of poorly behaving proteins*. J Biomol NMR, 2001. **20**(1): p. 11-4.
11. Gazzano-Santoro, H., et al., *Competition between rBPI23, a recombinant fragment of bactericidal/permeability-increasing protein, and lipopolysaccharide (LPS)-binding protein for binding to LPS and gram-negative bacteria*. Infect Immun, 1994. **62**(4): p. 1185-91.

Chapter 4

Interactions of lipopolysaccharide and polymyxin studied by NMR spectroscopy

4.0 Introduction

Antimicrobial peptides are evolutionarily ancient weapons to evade the harmful bacteria and they play a fundamental role for the successful evolution of complex multicellular organisms. Many of the host defense peptides, which have antimicrobial activity, are also able to bind to and neutralize LPS; however, these two activities do not necessarily correlate. In the light of occurrence of bacterial strains with multiple resistances against most antibiotics, antimicrobial peptides that interact with the outer layer of gram-negative bacteria, such as polymyxin (PMX), have recently received increased attention. Polymyxins a group of cyclic polycationic polypeptides with a long hydrophobic tail isolated from various strains of Bacillus polymyxa and related species[1]. They are selectively toxic for Gram-negative bacteria due to their specificity for the lipopolysaccharide molecule which forms the main constituent of the outer membrane of gram-negative bacteria.

LPS is known for its toxic effects. Septic shock is an overexaggerated host immune response to the presence of infection, particularly to LPS and is a significant cause of morbidity and mortality in all hospitalized patients. Polymyxin B (PMX-B) is one of the most efficient compounds for the treatment of septic shock, due to its ability to bind and detoxify bacterial lipopolysaccharide. Here we present a study of the interactions of PMX-B, -E and -M with lipopolysaccharide (LPS) from a deep-rough mutant strain of *E. coli*. A method for efficient purification of biosynthetically produced LPS using RP-HPLC in combination with ternary solvent mixtures was developed.

Polymyxin M (Mattacin) was produced biosynthetically from *Paenobacillus kobensis M*. LPS was incorporated into a membrane model, dodecylphophocholine (DPC) micelles, and its interaction with polymyxins was studied by heteronuclear NMR spectroscopy. Data from chemical shift mapping using isotope labelled LPS or labelled polymyxin, as well as from isotope-filtered NOESY experiments, highlight the mode of interaction of LPS with polymyxins. Using MD calculations in explicit water the complex of LPS with PMX-B in the presence of DPC micelles was modelled using restraints derived from chemical shift mapping data and intermolecular NOEs. In the modelled complex the macrocycle of PMX is centered around the phosphate group at GlcN-B, and additional contacts from polar side chains are formed to GlcN-A and Kdo-C, while hydrophobic side chains penetrate the acyl chain region.

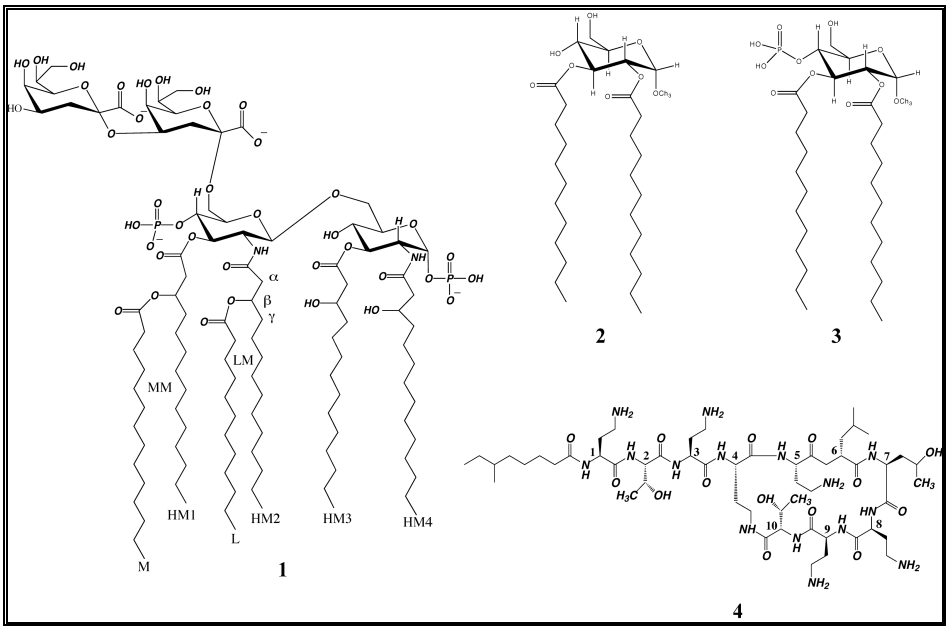


Figure 1: Chemical structures of LPS from the D31m4 *E. coli* strain (**1**), the two model compounds for LPS (**2** and **3**) and PMX-M (**4**). Note that in PMX-B residue D-Leu-6 is replaced by D-Phe and Thr-7 is replaced by Leu, and in polymyxin-E Thr-7 is replaced by Leu. In the text, the glycosidic residues of LPS are identified with the letters A-D, starting from the far right GlcN residue.

4.1 Experimental procedures

4.1.1 Materials

$^{15}NH_4Cl$ was purchased from Spectra isotopes (Columbia, USA), perdeuterated DPC-d38 (99%-d), and D_2O were ordered from Cambridge isotope laboratories (Andover, Massachusetts, USA). Methyl-5-doxylstearic acid was bought from Aldrich (Buchs, Switzerland). The Re-LPS producing strain D31m4 was purchased from the *E. coli* genetic resource center, New Haven. The PMX-M producing strain *Paenibacillus kobensis* M was obtained from Prof. J.C. Vederas and PMX-B and -E were purchased from Sigma Aldrich laboratories.

4.1.2 Production of ^{13}C-labelled LPS from the E. coli strain D31m4

Cells from the D31m4 strain of *E. coli* were grown at 37°C to an OD of around 1.0 at 600 nm on minimal medium M9 using 4 g ^{13}C glucose and 1 g ^{15}N ammonium chloride supplemented with 100 mg of Trp, His and Pro per liter. After harvest cells were resuspended in 50 ml of ice-cold water and pelleted down. To the pellet a minimum amount of cold water was added such that a thick paste was formed. LPS together with other components was precipitated through addition of 90 ml ice-cold methanol and centrifuged at 8000g. The pellet was resuspended in 90 ml ice-cold acetone, homogenized and centrifuged again, followed by another acetone washing step. The lyophilized cells (ca. 0.7 g) were taken up in 50 ml of a phenol:chloroform:petroleum ether (4:10:16; v:v) solvent mixture and centrifuged at 9200g, after which most of the LPS was contained in the supernatant. The remaining pellet was extracted once more to increase the yield in LPS. The supernatant was concentrated under a nitrogen stream and 2 ml of water were added dropwise to the concentrate. A waxy precipitate was formed followed by three cycles of washing with methanol and subsequent centrifugation. Thereafter the pellet was dried and lyophilized, after which it could only be resuspended/dissolved in water using repetitive additions of small amounts of water followed by sonication. Solubilization was improved upon adding aqueous 0.1M EDTA in the first portions. The resulting solution was centrifuged at 200000 g overnight and then lyophilized.

The lyophilized pellet after resuspension in mixture of solvent A (methanol:chloroform:water 57:12:31 v/v/v), and aqueous EDTA was directly loaded onto the HPLC column. For optimal purification of LPS a gradient system involving ternary solvent mixtures was used consisting of solvent A in 10mM NH_4Cl and methanol:chloroform (29.8:70.2 v/v, solvent B) in 50mM NH_4Cl. LPS (6 mg) was dispersed in a two phase system formed from 0.8 ml of solvent A and 0.2 ml of 0.1 M aqueous EDTA pH=7 and loaded directly onto the RP-C8 column. Chromatographic separation was achieved using the following gradient of solvents A and B: 2 CV of 2% B, 3 CV (2-17% B), 3.5 CV (17-27% B). UV detection was impossible and hence fractions were lyophilized and their content checked by MALDI-TOF using 6-aza-2-thiothymine as the matrix. Elution of the desired LPS occurred around 20-23% of solvent B.

4.1.3 Production of ^{13}C,^{15}N-labelled PMX-M from P. kobensis M

The producer strain, *Paenibacillus kobensis* M, was grown aerobically at 30°C on tryptic soy agar (TSA). A 1-liter batch of M9 medium was inoculated with a 10-ml *P. kobensis* M preculture (1% inoculum). After a total growth time of 16–24 h at 30°C with shaking (200 rpm), the cells were removed by centrifugation (1 hr, 10,000 rpm) and the supernatant was then passed through a Amberlite XAD-16 column. After washing with 30% ethanol, active peptide was then eluted with 70% isopropanol, which was adjusted to pH 2 with 12N HCl. All fractions were assessed for antimicrobial activity using a well plate assay. The contents of the active fraction were applied to a Superdex peptide 10/300 column (Amersham Biosciences). Fractions were collected for 3 column volumes with pure MilliQ Water and each assayed for activity. All active fractions were pooled, concentrated and applied as 20% isopropanol solutions to C18 reverse-phase HPLC. Complete purification required two separate steps of C18-HPLC. The first separation used a gradient of water/isopropanol gradient (0.1% trifluoroacetic acid), from 20% to 50% isopropanol, and the second step a water/methanol gradient (0.1% trifluoroacetic acid), from 45% to 85% methanol. PMX-M eluted at around 55%. Finally, 8 - 10 mg of pure PMX-M was obtained as slightly yellowish powder from a 1-liter culture and its chemical nature verified by ES1-MS (exp. mass: 1224.73Da, theoret. mass:1223.57 Da). During all steps of expression and purification antimicrobial activity was monitored by

inhibition of growth of an indicator strain. Agar plates were prepared by inoculating molten TSA (40 g/liter) with a culture of the indicator organism *E. coli* (1.0% inoculum). Small wells (app 4.6 mm diameter) were made in the seeded agar plates and 50-ml of filtered culture supernatant were added to the wells. Plates were incubated at 30°C, and the growth of the indicator organism was visible after approx.3 h[2].

4.1.4 NMR spectroscopy

LPS samples used in this work contained approx. 1mM LPS, 300 mM d_{38}-DPC in 40 mM d_{13}-MES D_2O buffer pD=5.8. Interaction studies with PMX were performed in acetate buffer in D_2O or H_2O/D_2O 9/1 pH=4.4. All spectra were recorded on Bruker AV-600 or AV-700 NMR spectrometer at T=310K. Proton and carbon chemical shifts were calibrated to DSS, and nitrogen shifts were referenced indirectly to liquid NH_3[3]. The spectra were processed using the Bruker Topspin2.0 software and transferred into CARA [4] or SPARKY[5] programs for further analysis.

For chemical shift assignments of $^{13}C,^{15}N$-labeled LPS 2D versions of 3D double- and triple-resonance experiments were recorded. In general, experiments used coherence selection schemes via pulsed-field gradients[4] and sensitivity-enhancement building blocks[6, 7] whenever possible. For assignments of the carbon spin systems in the lipid chains and the sugar units (H)CCH experiments recorded with 4 and 12 ms DIPSI-2 C-C mixing cycles were used. Linkage of the lipid chains onto the glucosamine parts of lipid A was achieved via correlations with the amide nitrogens using HNCA and HN(CO)CA experiments. To distinguish the two Kdo units key NOEs derived from a ^{13}C-resolved NOESY were exploited. Assignment of all resonances of polymyxin was done using HN(CO)CACB, HNCACB[8] and (H)CCH experiments[9, 10] analogous to the procedure used for proteins. Because of the small size of the peptide 2D versions were recorded with a total of less than 12 h measuring time for acquiring all spectra. Assignments of polymyxin-B and –E were based on assignments from PMX-M adjusted by using additional 2D heteronuclear spectra. In the spin label experiments, a 0.5 mM solution of LPS was separated into two aliquots, and to one of these 5-doxylstearate methyl ester was added so that the final concentration corresponded to approximately one

spin-label per micelle. Signal intensities from the two corresponding constant-time [^{13}C,^{1}H]-HSQC were extracted and the ratio of signal intensities from the samples with and without spin label was calculated.Measurements of interactions between LPS and PMX by chemical shift mapping observing LPS resonances utilized a 350 mM solution of ^{15}N,^{13}C Re-LPS and equimolar unlabelled peptides. Chemical shift changes in PMX-M were monitored in 200 mM solutions of ^{15}N,^{13}C -labeled PMX-M and equimolar unlabelled Re-LPS.

4.1.5 Molecular dynamics calculations

All calculations were preformed within the program GROMACS [11]. Briefly, coordinates of LPS were adapted from the pdb entry 1QFF and coordinates of polymyxin B were built using the program Ghemical[12]. Parameters and topologies of PMX-B and LPS as well as partial charges of PMX-B for GROMACS were established based on data from PRODRG server[13] and the GROMOS 53a6 forcefield[14]. Partial charges of LPS were assigned for a protonation state corresponding to experimental conditions of pH by the MPEOP method[15, 16]. Parameters for DPC were derived from values for DPPC from the GROMOS force field library.

A detailed description of the methodology persued can be found in the Supp. Mat. Briefly, the system was prepared as follows: i) An initial complex between LPS and polymyxin B was prepared and equilibrated in the presence of a DPC micelle. ii) A set of simulated annealing calculations was performed yielding 450 structures. The DPC molecules were explicitly included in the system, water molecules were substituted by implicit solvent and the dominant NOE-derived upper distance limits were included (force constant was set to 1000 [kJ mol^{-1} nm^{-2}]). iii) Those structures were selected, which displayed best agreement with the chemical shift mapping data. iv)These were then equilibrated with explicit solvent and subjected to further refinement and analysis. The latter included an assesment of the stability of the MD trajectory and a comparison of the average intermolecular distances with chemical shift mapping data.

4.2 Results

4.2.1 Production and purification of ^{13}C-labeled LPS

E. coli strains from the deep rough *E. coli* mutant D31m4 were isolated from the membrane fraction of cells grown on minimal medium containing ^{13}C-glucose and ^{15}N-NH_4Cl as the sole carbon and nitrogen sources, respectively. After PCP extraction and further purification using published protocols[17] the yield was ca. 129 mg/l of culture. Remaining impurities were removed by RP-HPLC using a ternary solvent mixture. In this procedure the solvent system was carefully adapted to form a single phase over the whole gradient range of solvent A (methanol: chloroform:water) and solvent B (methanol:chloroform) at RT. Importantly, any mixture of these two solvent systems is relatively close to a two-phase system, and this condition proved to have favorable properties for dissolving LPS. MALDI-TOF spectra of LPS before and after this HPLC purification step are depicted in Fig. 2. Sufficient quantities (40 mg) of chemically pure LPS for the NMR studies could be produced from 1 L of culture using this protocol. As demonstrated in Fig. S4 in the Supp. Mat. this method is capable of separating pyrophosphate from the monophosphate derivatives.

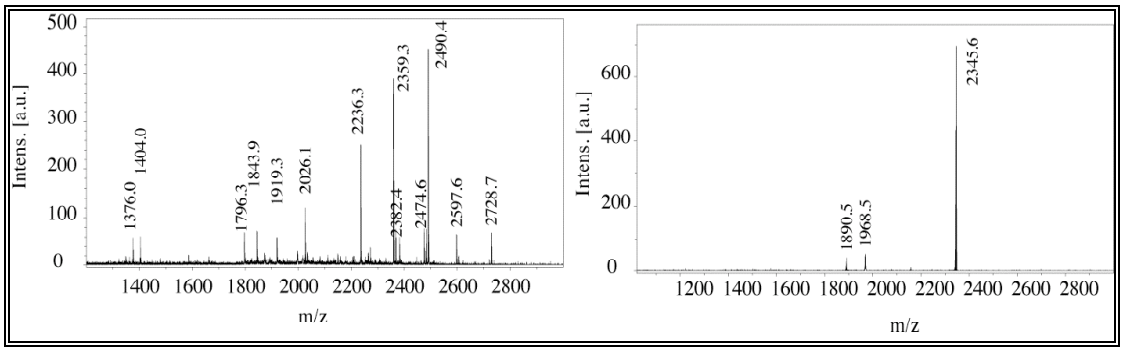


Figure 2: Comparison of the MALDI-TOFF MS spectrum of commercial LPS with LPS from the deep rough D31m4 *E. coli* strain purified by the protocol that includes an additional HPLC purification step.

4.2.2 Assignment of LPS and polymyxin resonances

Chemical shift mapping[18] or NOE-based methods[19] can be used to study biomolecular interactions (see also[20-22]). Both methods potentially deliver information on interacting moieties but require assignments of chemical shifts. The best chemical shift dispersion is usually available in heteronuclear shift correlation spectra (e.g. [^{15}N,^{1}H]-HSQC or [^{13}C,^{1}H]-HSQC spectra). Importantly, these experiments still work well in the presence of the increased line widths that are usually present in systems that are stably anchored into phospholipids micelles. In addition, as was unfortunately the case in some of our applications, additional exchange broadening occurred upon complex formation. To probe integration of LPS into DPC micelles and to study its interaction with peptides we decided to label it with ^{13}C and ^{15}N isotopes and use the corresponding HSQC spectra for chemical shift mapping.

To assign all signals in the constant-time [^{13}C,^{1}H]-HSQC spectrum[23], 3D (H)CCH-TOCSY spectra[9, 10] served for assignment of spin systems (Fig. 3). The C,H-plane of the HNCA[6, 24] and HN(CO)CA[24] experiments were used to establish scalar connectivities between terminal carbons of the fatty acid chains and C-2 of the glucosamine moieties. Hydroxymyristoyl (HM) can be distinguished from lauroxymyristoyl (LM) and thereby helps to differentiate between GlcN-A from GlcN-B (see legend of Fig. 1). Due to chemical shift degeneracies it was impossible to assign chains of the fatty acids (myristoyl of myristoxy-myristoyl and lauroyl of lauroxymyristoyl). The two Kdo units were linked and thereby distinguished from each other using several key NOEs in the 3D ^{13}C-NOESY spectra.

The unique chemical shifts of the C3 moiety of Kdo is located in a region separated from all other sugar resonances, and was used as a starting point for sequential assignment. Only the methylene group C3 of Kdo-C is expected to receive an NOE from H(C6) of GlcN-B. This assignment was additionally supported by the fact that H(C6) and H(C7) of Kdo-C displayed an NOE to H(C3) of both Kdo units, which is unlikely to be the case for H(C6) of Kdo-D. The ^{1}H,^{13}C and ^{15}N chemical shifts of LPS in DPC micelles are reported in Table 1.

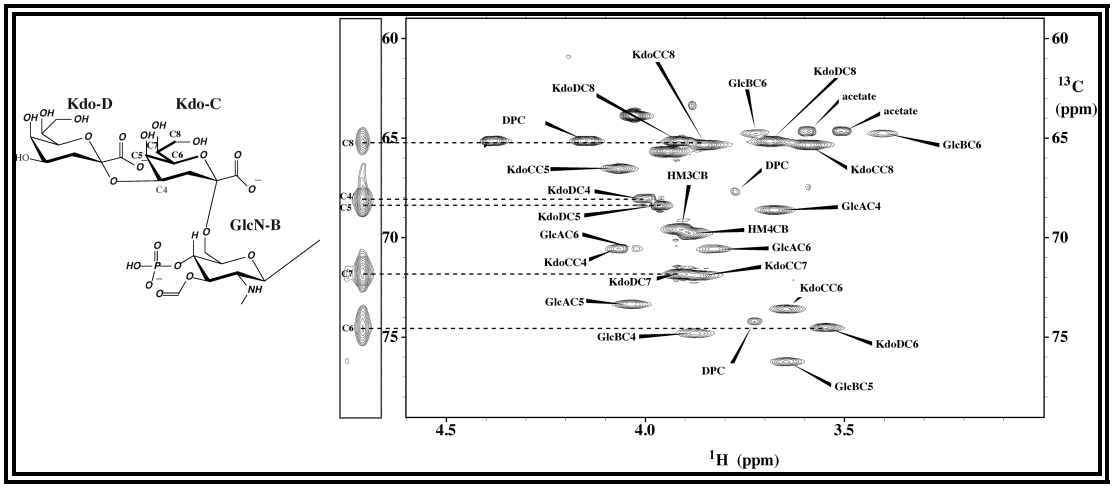


Figure 3: Assignment of ^{13}C spin systems of a short stretch contained in Kdo-D from LPS using (H)CCH-TOCSY spectra. On the right the constant-time [^{13}C,^{1}H]-HSQC spectrum is shown, in the middle a strip from the (H)CCH TOCSY taking at the proton frequency of H-(C4) of Kdo-D.

4.2.3 Topology of LPS and polymyxin insertion into the DPC micelle

To probe whether LPS properly inserts into the DPC micelles, and whether the sugar moieties really protrude into the aqueous phase, the micelle-integrating spin label methyl 5-doxylstearate was used. The paramagnetic moiety of the spin label resides in the headgroup region[25, 26], and signal reductions in the proton-carbon correlation maps indicate proximity to the water-micelle interface. The results from the spin-label experiment are depicted in Fig. 4. Strongest attenuations of signals are observed for C,H moieties of terminal acyl chain carbons, that are close to the GlcN sugars. Much less pronounced attenuations are observed for signals from the GlcN moieties, and the Kdo units are essentially not affected. The data demonstrate that LPS inserts into the DPC micelles such that the amide moieties of LPS are located in the headgroup region, the acyl chains are inserted into the micelle interior, and the carbohydrate units are exposed on the surface of the micelles. We additionally probed for the topology of the polymyxin-micelle complex using the micelle-integrating spin-label methyl 5-doxylstearate as described in the previous section. Notably PMX-B, -E and -M differ only for residues in position 6 and 7 (see legend of Fig.1). Largest attenuations were observed for the amide moieties of the first residue, and for residues 6 and 7, the latter two present hydrophobic

or aromatic residues (Table S11), revealing that polymyxins are attached via the N-terminal lipid chain and the apolar residues to the micelle.

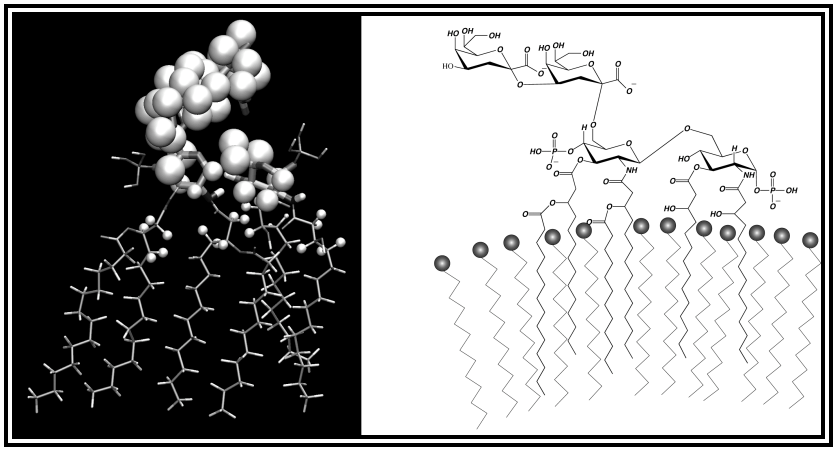


Figure 4: Left: Results from the spin-label experiments displayed on the structure of LPS. The size of the sphere is proportional to the remaining signal in the proton-carbon correlation map. Because signals from ends of the fatty acid chains are not resolved in the spectra, no encoding is shown for these atoms. Right: Sketch of LPS from the D31m4 strain when inserted into DPC micelles. The phospholipid headgroups are depicted as grey spheres.

4.2.4 Interactions of LPS with various types of polymyxins

Studies of interactions of LPS with polymyxin were performed using chemical shift mapping and isotope-filtered NOESY experiments. We initially used very simplified model compounds of LPS derived from a-(D)-glucose, that lack Kdo and the second glucosamine units (Fig.1). Considering the proposed importance of phosphate moieties for binding it was not surprising that compound **3** did not display any interactions with PMX-B, but also compound **4**, that contains the presumably important phosphate group, did not cause any changes in the spectra, indicating that the Kd is (much) larger than 1mM.

The [^{15}N,^{1}H]- and the [^{13}C,^{1}H]-HSQC spectra of PMX-M revealed large changes (*vide infra*). The chemical shift changes of LPS carbon and proton frequencies upon adding PMX-B,-E or -M are mapped onto the structure in Fig. 5. In general, the largest chemical shift changes of LPS resonances were observed for all observable atoms of

GlcN-B. Interestingly, resonances from C1 and C2 of GlcN-A were affected only very little (Fig. S3A and S3C). In addition, resonances from Kdo-C display significantly larger changes than those of Kdo-D, with C-3 of Kdo-C being shifted by the largest extent (see Fig. S3F). Moreover, large changes in chemical shift are observed for the alpha (Fig. S3G), beta and gamma (Fig. S3H) positions of the fatty acid chains, in particular for those of HM2 (note that the second branch of MM and LM cannot be assigned due to spectral overlap). The Ca change for HM1 is less pronounced. The alpha position of HM4 was apparently more affected than of HM3. The chemical shifts changes on the beta position were smaller and clear only for HM4.

A comparison of chemical shift changes of LPS between its complexes with PMX-B, PMX-E and PMX-M may serve to identify differences in their binding modes. Chemical shift changes are almost identical at positions C-4, C-5 of GlcN-B as well as on position C-3 of Kdo-C for PMX-B and PMX-E (Fig. S3E and S3F), and very similar changes occur for the alpha position of HM2. We suspect that these regions are in contact with the conserved parts of polymyxins. However, significant differences were observed for C-3 of GlcN-B and moderate differences can be seen at positions of C-3, C-5 and C-6 of GlcN-A and at position C-1 of GlcN-B (Fig. S3B). Addition of PMX-M results in line-broadening for resonances of C-4 and C-5 of GlcN-B, as well as for C-3 of Kdo-C, and in a chemical shift difference for C-6 of Kdo-C. We address this observation to a slightly different binding mode of PMX-M, which results in different contacts to Kdo-C.

Interactions between LPS and polymyxin were additionally detected in heteronuclear spectra when $^{13}C,^{15}N$-labelled PMX-M was mixed with unlabeled LPS. Unfortunately, very broad lines and multiple sets of resonances were detected in the [$^{15}N,^{1}H$]-HSQC of polymyxin-M in presence of LPS indicating the presence of exchange processes. Moreover, the fact that the signals were shifted considerably and, even more importantly, the presence of very broad line precludes detailed assignment. Even when changing pH or adding other detergents such as SDS we have not been able to achieve conditions that result in spectra of sufficient quality. Considering that interactions between LPS and polymyxin are suggested to be electrostatic in nature[27], we decided

to reduce the strength of the interaction by adding small amounts of divalent metal ions. When adding PMX-B, -M or –E in the presence of Ca^{2+} or Mg^{2+} we noticed slight precipitation. In the presence of 20 mM $MgCl_2$ a good-quality [^{15}N,^{1}H]-HSQC was obtained, and the peak positions are close to those in the absence of LPS. Fortunately, linewidths in the [^{13}C,^{1}H]-HSQC spectra are smaller such that interactions can be directly detected by shift mapping methods.

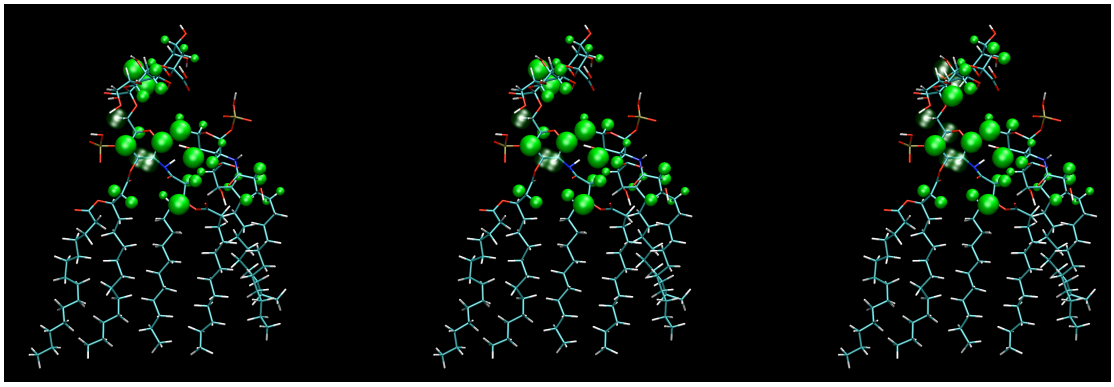


Figure 5: Changes in ^{1}H and ^{13}C chemical shifts of resonances from LPS when adding PMX-B (A), PMX-E(B) and PMX-M(C). The size of the spheres is proportional to the observed differences. Deviations of peak positions were extracted from the [^{13}C,^{1}H]-HSQC spectra and computed according to $\Delta\delta=\mathrm{sqrt}((\Delta\delta\ C)^2+(10X\Delta\delta\ ^{1}H)^2$. Those atoms that are exchange-broadened beyond detection are depicted by green transparent spheres

The results are summarized in Table S3, and shall only be briefly summarized here: largest chemical shift changes occur for Ca resonances of residues 4, 5, 6, 8 and 9 of polymyxin-M, the members of the macrocyclic ring. Interestingly, smaller differences were observed for resonances from the lipid chain or from Dab-1, Thr-2 or Dab-3. The fact that only residues from the macrocycle experience large changes, and the observation, that the changes in the macrocycle occur at certain but not in all positions, indicates that the heptapeptide ring of PMX binds in a rather well-defined manner to LPS. In addition to chemical shift mapping we have also conducted isotope-filtered NOESY experiments, in which ^{13}C-labeled LPS was mixed with unlabelled PMX-B, -E, so that the spectra only contain intermolecular NOEs (see Table 2). Interestingly, in the case of PMX-B, NOEs between side chain protons of Phe-6 or Leu-7 and the lipid chains

of LPS are detected, supporting the results from the shift mapping data. In particular, a number of NOEs between protons of the p-system of Phe-6 and the lipid chains are observed, in particular to Cg and Ca of HM3. In the case of PMX-M position 7 is occupied by the polar Thr residue. The fact that PMX-B or PMX-E form more contacts with LPS that involve side chains of residues 6 and 7 indicates that the hydrophobic nature of this side chain may be important for orienting polymyxin in the complex with LPS.

To obtain a first picture of the complex formed between LPS and PMX-B we have performed a MD calculation in which LPS was docked to PMX-B using NOE-derived upper distance limits (Fig. 6) in explicit water and in the presence of a DPC micelle. In the complex backbone amide moieties of the peptide macro-cycle form contacts with the phosphate group of GlcN-B. Residues from one hydrophobic site - D-Phe-6 and Leu-7 - are in proximity to the fatty acid chains of LPS. Leu-7 makes additional contacts with the GlcN-B moiety. The amino group of Dab-8 is involved in electrostatic contacts with the phosphate group at GlcN-A. Amino groups of Dab-9 and -3 are possibly forming electrostatic interactions with the carboxyl group of Kdo-C.

The backbone of residues Dab-1 to Dab-3 is located in the acyl region of the branched fatty acids originating from GlcN-B. To summarize most contacts in this model structure are made with GlcN-B and the adjacent atoms of GlcN-A, as well as with parts of Kdo-C, and the model structure is supported by the chemical shift mapping data. In addition, interacting moieties of LPS that make contacts common with all polymyxin variants are in contact with conserved parts of these peptides (residues 1 to 4 and 8 to 10), while those sites that differ in their interaction with the different peptides form contacts with non-conserved residues.

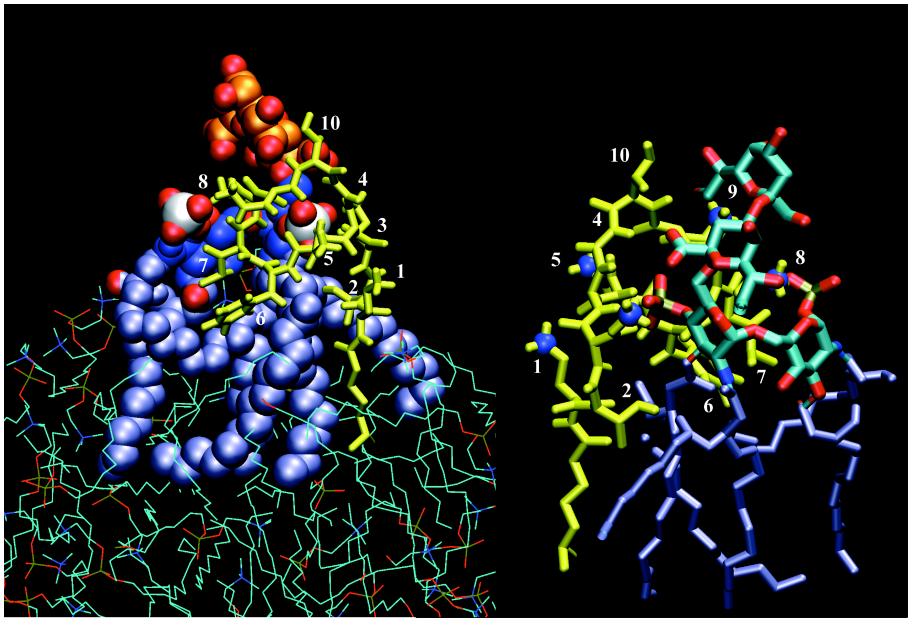


Figure 6: Left: Structure of the PMX-B—LPS complex as derived from the MD calculation. The covalent structure of PMX-B is depicted in yellow. Heavy atoms of LPS are drawn as van der Waals spheres, with lipid chains colored in ice-blue, GlcN carbons in dark blue, Kdo carbons in orange, and phosphorous atoms in white. DPC molecules are indicated by thin lines. Residue numbers of PMX-B are placed close to the corresponding sidechains. Right: Representation showing only bonds of LPS and PMX-B. PMX-B is again drawn in yellow, with amino nitrogens of DAB residues in blue, while LPS bonds are drawn in ice-blue (lipids) or green (sugar parts).

4.3 Discussion

Sepsis caused by gram-negative bacteria is a serious source of mortality in many clinical cases, accounting for approximately 200,000 deaths in the US annually (see David[28] and references therein). The primary trigger for sepsis was identified as endotoxin, and endotoxin-neutralizing agents are therefore valuable therapeutics. PMX-B is considered as the "gold-standard" for LPS-sequestering agents. It is a highly cationic decapeptide containing six diaminobutyric acid (Dab) residues, a macrocylic ring involving residues 4 to 10, and an acyl chain coupled to the N terminus. In this work we have used polymyxins of different types to study interactions with LPS from the deep rough mutant *E. coli* strain. To our knowledge this work for the first time presents experimental data on the interacting moieties in a membrane mimetic environment.The conformation of LPS-bound PMX-B and -E was elucidated previously using transfer-

NOE techniques by Pristovsek. [29] Therein, the LPS-bound peptides assume envelope-like bent cycles that separate the two hydrophobic residues 6 and 7 from the charged Dab residues 4, 5, 8 and 9. The conformation from the transfer-NOE derived structure was then docked onto a lipid A model. In the resulting complex the lipid chain forms transient contacts with acyl chains A and B from lipid A. Moreover, the two phosphate groups are in close contact with the g-NH_2 groups of Dab residues 1/5 and 8/9. The data highlight the importance of electrostatic interactions between phosphate groups and amino groups of polymyxin as well as hydrophobic interactions of the lipid chains with Phe or Leu residues. Martin *et al.* have studied the structure of LPS-bound PMX-M [2], which displayed a chair-like confor-mation, in which the side chain of Dab-4 or -8 and Leu-6 or Thr-7 point into the opposite directions in a fashion similar to that proposed for PMX-B and –E [2]. Pristovsek determined the structure of a LPS-bound synthetic fragment of the LALF protein [30] revealing a hairpin-type fold again characterized by spatial separation of hydrophobic and cationic residues. In another study Bhunia *et al.* investigated binding of melittin to LPS micelles [31]. Cationic residues have also been postulated to critically contribute to binding of LPS to proteins such as MD-2 [32] or FhuA [33, 34].

Herein we set out to determine interactions between LPS and polymyxins in more detail. Considering that interacting chemical moieties have not been experimentally identified with confidence so far we reasoned that these studies should use chemically defined environments in a setup that mimics natural conditions. To achieve chemical homogeneity we used biosynthetic LPS to facilitate isotope labeling. LPS from the deep rough mutant of *E. coli* presents a simple system containing all chemical moieties believed to be important for the interaction. LPS in gram-negative bacteria is embedded in a phospholipid bilayer, and therefore the system was studied in DPC micelles. It could be shown that LPS properly inserts into the micelle via integration of the acyl chains into the micelle interior, and the GlcN and Kdo portions are fully exposed in the aqueous compartment. LPS from the same *E. coli* strain has been assigned previously using homonuclear 2D NMR techniques.[35, 36]. For the purpose of the study the development of chromatographic techniques for LPS was very important that allow studying the unmodified LPS. Using ^{13}C isotope labeling we have been able to fully

assign LPS while integrated into DPC micelles. In the latter environment spectra with reasonable line widths and resolution can be recorded, allowing the study of interactions with polymyxin in sufficient detail. Our experimental data now confirm the model of the PMX-B LPS complex, proposed by Pristosek et al. [29] In their complex the macrocycle of PMX-B covers the GlcN disaccharide unit, and hydrophobic sidechains of polymyxin form contacts with the lipid chains from LPS, and a similar complex topology was proposed for the interaction of LPS with PMX-M [2]. Such an arrangement requires a reasonable separation of hydrophobic and hydrophilic side chains in polymyxin, so that hydrophobicity matching between polymyxin and LPS can help to orient it. It should be noticed that LPS in its natural environment is embedded in a phospholipid environment. Hydrophobicity matching of moieties of polymyxin with the corresponding parts of LPS in that environment is amplified, because similar requirements influence interactions of polymyxin with the phospholipids.

It is therefore of little surprise that Leu or Phe residues, for which favorable energies for partitioning into the membrane interior or the water-membrane interfacial region have been measured [37], form contacts with the a, b, and g acyl chain carbons of LPS. In PMX-M position 7 is occupied by Thr, a much more polar residue. According to the data from this work, PMX-M may be slightly differently oriented, possibly forming some-what stronger contacts to the Kdo units. In the PMX-M—LPS complex electrostatic or polar interactions involving the Dab residues of polymyxin and the carbohydrate moieties of GlcN-B or Kdo-C and to a smaller extent, Kdo-D dominate, while in the case of PMX–B or –E additional hydrophobic interactions from sidechains of Leu-7 are likely to contribute to binding. While our work confirmed the contacts made with moieties of the GlcN units proposed in the model complex from Pristovsek [29] it is different in that it emphasizes the importance of contacts to Kdo units, in particular to Kdo-C.

Different models have been proposed for the mechanism of action of polymyxin. Shai et al. have proposed that binding of polymyxin to the core part of LPS, lipid A, results in disturbance of the LPS-phospholipid bilayer destroying the integrity of the

outer membrane, and possibly leading to pore formation [38, 39]. The orientation of LPS in the phospholipid micelles as measured using spin labels has demonstrated that the carbonyl group region of the LPS lipid chains is located in the headgroup region. Binding of polymyxin to LPS therefore places the amphiphilic peptide in a similar position compared to binding to pure phospholipid micelles. Accordingly, similar mechanisms for membrane permeabilization are plausible.

4.4 Acknowledgements

We are very grateful to John C. Vederas for supplying us with cells producing polymyxin M and John Robinson for helpful discussions. We would like to thank for financial support from the Swiss National Science Foundation (grant No. 3100A0-11173).

4.5 References

1. Dixon, R.A. and I. Chopra, *Polymyxin B and polymyxin B nonapeptide alter cytoplasmic membrane permeability in Escherichia coli.* J Antimicrob Chemother, 1986. **18**(5): p. 557-63.
2. Martin, N., et al., *Isolation, structural characterization, and properties of mattacin (Polymyxin M), a cyclic peptide antibiotic produced by Paenibacillus kobensis M.* J Biol Chem, 2003. **278**(15): p. 13124-13132.
3. Live, D.H., et al., *Observation of 1000-fold Enhancement of ^{15N} NMR via Proton-Detected Multiple-Quantum Coherences: Studies of Large Peptides.* J. Am. Chem. Soc., 1984. **106**: p. 6104-6105.
4. Keller, R., *The computer aided resonance assignment*. 2004, Goldau: CANTINA Verlag.
5. Goddard, T.D. and D.G. Kneller, *SPARKY 3*, University of California, San Francisco
6. Kay, L.E., et al., *Three-Dimensional Triple-Resonance NMR Spectroscopy of Isotopically Enriched Proteins.* J. Magn. Reson., 1990. **89**: p. 496.
7. Palmer, A.G., et al., *Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation NMR spectroscopy.* J. Magn. Reson., 1991. **93**: p. 151-170.
8. Wittekind, M. and L. Mueller, *HNCACB, a High-Sensitivity 3D NMR Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha-Carbon and Beta-Carbon Resonances in Proteins.* J. Magn. Reson. Ser. B, 1993. **101**(2): p. 201-205.
9. Olejniczak, E.T., R.X. Xu, and S.W. Fesik, *A 4D HCCH-TOCSY experiment for assigning the side chain 1H and 13C resonances of proteins.* J Biomol NMR, 1992. **2**(6): p. 655-9.

10. Bax, A., et al., *Practical Aspects of Proton-Carbon-Carbon-Proton Three-Dimensional Correlation Spectroscopy of 13C labelled Proteins*. J. Magn. Reson., 1990. **87**: p. 620-627.
11. Van, d.S., D, et al., *GROMACS: Fast, flexible, and free.* J Comput Chem, 2005. **26**(16): p. 1701-1718.
12. Hassinen, T. and M. Perakyla, *New energy terms for reduced protein models implemented in an off-lattice force field.* J Comput Chem, 2001. **22**(12): p. 1229-1242.
13. Schüttelkopf, A.W. and D.M.F. van Aalten, *PRODRG – a tool for high-throughput crystallography of protein-ligand complexes*. Acta Crystallogr. D Biol. Crystallogr., 2004. **60**: p. 1355-1363.
14. Oostenbrink, C., et al., *A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6.* Journal of Computational Chemistry, 2004. **25**(13): p. 1656-1676 %U http://dx.doi.org/10.1002/jcc.20090.
15. No, K.T., J.A. Grant, and H.A. Scheraga, *Determination of net atomic charges using a modified partial equalization of orbital electronegativity method. 1. Application to neutral molecules as models for polypeptides.* The Journal of Physical Chemistry, 1990. **94**(11): p. 4732-4739 %U http://pubs.acs.org/doi/abs/10.1021/j100374a066.
16. Gasteiger, J. and M. Marsili, *Iterative partial equalization of orbital electronegativity--a rapid access to atomic charges.* Tetrahedron, 1980. **36**(22): p. 3219-3228 %U http://www.sciencedirect.com/science/article/B6THR-42P5V46-14/2/fc6272b08de03d3145a09adaae190f00.
17. Galanos, C., O. Luderitz, and O. Westphal, *A new method for extraction of R-Lipopolysaccharides* Eur. J. Biochem., 1969. **9**: p. 245-249.
18. Shuker, S., et al., *Discovering high-affinity ligands for proteins: SAR by NMR.* Science, 1996. **274**(5292): p. 1531-4.
19. Clore, G.M. and A.M. Gronenborn, *Theory and Applications of the Transferred Nuclear Overhauser Effect to the Study of the Conformations of Small Ligands Bound to Proteins.* J. Magn. Reson., 1982. **48**: p. 402-417.
20. Gemmecker, G., *Isotope Filter and Editing Techniques*, in *BioNMR in Drug Research*, O. Zerbe, Editor. 2003, Wiley-VCH: Weinheim. p. 373-390.
21. Blommers, M.J.J. and S. Rüdisser, *NMR of Weakly Binding Ligands*, in *BioNMR in Drug Research*, O. Zerbe, Editor. 2003, Wiley-VCH: Weinheim. p. 355-372.
22. Marchioro, C., et al., *Experiments in NMR-Based Screening*, in *BioNMR in Drug Research*, O. Zerbe, Editor. 2003, Wiley-VCH: Weinheim. p. 321-340.
23. Vuister, G.W. and A. Bax, *Resolution enhancement and spectral editing of uniformly C-13-enriched proteins by homonuclear broad-band C-13 decoupling.* J. Magn. Reson., 1992. **98**: p. 428-435.
24. Grzesiek, S. and A. Bax, *Improved 3D triple-resonance NMR techniques applied to a 31-kDa protein* J. Magn. Reson., 1992. **96**: p. 432-440.
25. Jarvet, J., et al., *Three-dimensional structure and position of porcine motilin in sodium dodecyl sulfate micelles determined by ^{1H} NMR.* Biochemistry, 1997. **36**: p. 8153-63.

26. Papavoine, C.H., et al., *Location of M13 Coat protein in sodium dodecyl micelles as determined by NMR.* Biochemistry, 1994. **33**: p. 12990-12997.
27. Koch, P., et al., *Thermodynamics and structural studies of the interaction of Polymyxin B with deep rough mutant lipopolysaccharides.* J Colloid Interf Sci, 1999. **213**(2): p. 557-564.
28. David, S., *Towards a rational development of anti-endotoxin agents: novel approaches to sequestration of bacterial endotoxins with small molecules.* J. Mol. Recognit., 2001. **14**: p. 370-387.
29. Pristovsek, P. and J. Kidric, *Solution structure of polymyxins B and E and effect of binding to lipopolysaccharide: An MMR and molecular modeling study.* J Med Chem, 1999. **42**(22): p. 4604-4613.
30. Pristovsek, P., et al., *Structure of a synthetic fragment of the LALF protein when bound to lipopolysaccharide.* J Med Chem, 2005. **48**(5): p. 1666-1670.
31. Bhunia, A., P.N. Domadia, and S. Bhattacharjya, *Structural and thermodynamic analyses of the interaction between melittin and lipopolysaccharide.* Biochimica et biophysica acta, 2007: p. .
32. Gruber, A., et al., *Structural model of MD-2 and functional role of its basic amino acid clusters involved in cellular lipopolysaccharide recognition.* J Biol Chem, 2004. **279**(27): p. 28475-28482.
33. Ferguson, A., et al., *Siderophore-mediated iron transport: Crystal structure of FhuA with bound lipopolysaccharide.* Science, 1998. **282**(5397): p. 2215-2220.
34. Ferguson, A., et al., *A conserved structural motif for lipopolysaccharide recognition by procaryotic and eucaryotic proteins.* Structure, 2000. **8**(6): p. 585-592.
35. Qureshi, N., et al., *Complete structural determination of lipopolysaccharide obtained from deep rough mutant of escherichia-coli - purification by high-performance liquid-chromatography and direct analysis by plasma desorption mass-spectrometry.* J Biol Chem, 1988. **263**(24): p. 11971-11976.
36. Agrawal, A.B., Qureshi, Takayama, *1H and 13C NMR assignments of a lipopolysaccharide obtained from the deep rough mutant of Escherichia coliD31m4* Magn. Reson. Chem., 1998. **36**: p. 1-7.
37. Wimley, W.C. and S.H. White, *Experimentally determined hydrophobicity scale for proteins at membrane interfaces.* Nature Struct. Biol., 1996. **3**: p. 842-848.
38. Rosenfeld, Y., N. Papo, and Y. Shai, *Endotoxin (lipopolysaccharide) neutralization by innate immunity host-defense peptides - Peptide properties and plausible modes of action.* J Biol Chem, 2006. **281**(3): p. 1636-1643.
39. Papo, N. and Y. Shai, *A molecular mechanism for lipopolysaccharide protection of gram-negative bacteria from antimicrobial peptides.* J Biol Chem, 2005. **280**(11): p. 10378-10387.

Tables

Table 1: Carbon and proton chemical shifts of LPS from the deep-rough mutant of *E. coli*, 1mM LPS in 300mM DPC, pH= 5.8; 40mM MES, T=310K

Sugar moieties				
	GlcN-A	GlcN-B	Kdo-C	Kdo-D
H 1	5.346	4.636		
H 2	3.912	3.854		
H 3	5.136	5.066	1.879	2.079
H 3			1.929	1.707
H 4	3.675	3.872	4.066	4.000
H 5	4.037	3.644	4.067	3.961
H 6	4.022	3.403	3.644	3.547
H 6	3.828	3.722		
H 7			3.866	3.909
H 8			3.848	3.913
H 8			3.591	3.687
C 1	95.56	104.41		
C 2	54.07	55.57		
C 3	75.42	76.07	35.32	36.61
C 4	68.62	74.81	70.57	68.04
C 5	73.36	76.23	66.51	68.39
C 6	70.53	64.76	73.60	74.53
C 7			71.90	71.83

Lipid chains						
	LM		MM			
	L	HM2	M	HM1	HM-4	HM-3
HA	2.306*	2.562	*	2.643	2.312	2.383
HA		2.585	*	2.670	2.366	2.433
HB	1.563*	5.283	*	5.128	3.890	3.910
HC	1.210*	1.519	*	1.548	1.303	1.383
HC		1.559	*	1.601	1.380	1.471

HD					1.281	1.392
CA	36.51*	43.15	*	41.10	45.38	44.18
CB	27.34*	72.37	*	72.50	69.77	69.61
CG	31.50*	37.18	*	36.25	38.30	40.30
CD					27.98	28.08

* Resonances that cannot be distinguished between L and M because of overlap

Table 2: Intermolecular NOEs as detected in ^{13}C-filtered NOESY experiments performed with ^{13}C-labelled LPS and unlabelled PMX-B and –E.

Polymyxin-B						
HM4-H(Cα)	Leu7-Hδ1	w		HM3-H(Cγ)	Phe6-Hε	w
HM4-H(Cβ)	Leu7-Hδ1	s		HM3-H(Cδ)	Phe6-Hδ	s
HM4-H(Cδ)	Leu7-Hδ1	w		HM3-H(Cδ)	Phe6-Hε	s
HM4-H(Cδ)	Leu7-Hδ2	s		HM4-H(Cδ)	Phe6-Hδ	s
HM4-H(Cγ)	Leu7-Hδ2	s		HM4-H(Cδ)	Phe6-Hε	w
HM4-H(Cβ)	Leu7-Hδ2	s		HM1-H(Cβ)	Phe6-Hδ	w
HM4-H(Cα)	Leu7-Hδ2	s		HM4-H(Cβ)	Phe6-Hδ	w
HM4-H(Cβ)	Leu7-Hγ	s		HM3-H(Cγ)	Phe6-Hδ	s
HM4-H(Cβ)	Leu7-Hβ2	s		HM4-H(Cα)	Phe6-Hδ	s
HM4-H(Cβ)	Leu7-Hβ1	w		HM4-H(Cγ)	Phe6-Hδ	s
HM3-H(Cβ)	Phe6-Hβ2	w		HM3-H(Cα)	Phe6-Hδ	s
HM1-H(Cβ)	Phe6-Hβ2	w		HM3-H(Cβ)	Phe6-Hδ	s
HM2-H(Cα)	Phe6-Hβ1	w		HM2-H(Cα)	Phe6-Hδ	s
HM4-H(Cα)	DABA8-Hγ	s		HM2-H(Cβ)	Phe6-Hε	w
HM4-H(Cβ)	DABA8-Hγ	w		HM4-H(Cα)	DABA8-HN	w
HM4-H(Cγ)	Phe6-Hε	w				
Polymyxin-E						
HM4-H(Cα)	Leu6-Hδ1	m		HM2-H(Cα)	Leu6-Hδ1	s

HM4-H(Cα)	Leu7-Hβ	m		HM2-H(Cα)	Leu6-HN	m
HM4-H(Cβ)	Leu6-Hγ	m		HM2-H(Cα)	Leu6-Hγ	m
HM4-H(Cβ)	Leu7-Hβ	m		HM2-H(Cβ)	Leu6-Hγ	m
HM4-H(Cβ)	Leu7-Hδ1	m		HM1-H(Cγ)	Dab4-HN	w
HM4-H(Cβ)	Leu7-Hγ	w		HM1-H(Cγ)	Leu6-HN	m
HM4-H(Cβ)	Leu7-HN	m		HM3-H(Cα)	Leu6-Hδ1	m
HM4-H(Cδ)	Leu7-Hβ	w		HM3-H(Cβ)	Leu6-Hδ1	m
HM4-H(Cγ)	Leu6-Hδ1	s		HM3-H(Cγ)	Leu6-Hδ1	m

4.6 Supplementary materials

Table S1: Residual intensity of C,H cross peaks of LPS after addition of 4mM Methyl-5-doxylstearate (pH=4.4, 300mM DPC, 40mM Acetate, T 310K)

Unit	Atoms	Int/Int(0)	Unit	Atoms	Int/Int(0)
HM3	Cγ-Hγ(2)	0.36	GlcN A	C2-H2	0.79
HM3	Cγ-Hγ(1)	0.37	GlcN A	C1-H1	0.85
HM1	Cγ-Hγ(2)	0.45	GlcN A	C3-H3	0.87
HM1	Cγ-Hγ(1)	0.45	GlcN A	C6-H6	0.87
HM3	Cα-Hα(1)	0.46	Kdo C	C6-H6	0.89
HM3	Cβ-Hβ	0.46	Kdo C	C8-H8(2)	0.89
HM3	Cδ-Hδ	0.47	Kdo D	C7-H7	0.89
HM2	Cβ-Hβ	0.47	GlcN A	C6-H6	0.90
HM1	Cα-Hα(2)	0.47	Kdo D	C8-H8(1)	0.90
HM4	Cγ-Hγ(1)	0.48	Kdo C	C3-H3(1)	0.90
HM4	Cγ-Hγ(2)	0.50	GlcN B	C4-H4	0.90
HM3	Cα-Hα(2)	0.50	Kdo D	C8-H8(2)	0.90
HM2	Cα-Hα(1)	0.53	Kdo C	C8-H8(1)	0.91
HM2	Cα-Hα(2)	0.53	Kdo C	C4-H4	0.92
HM2	Cγ-Hγ(1)	0.54	GlcN B	C6-H6	0.92
HM4	Cδ-Hδ	0.55	Kdo D	C5-H5	0.92
HM1	Cβ-Hβ	0.55	Kdo C	C5-H5	0.92
HM2	Cγ-Hγ(2)	0.56	GlcN B	C5-H5	0.93
HM4	Cα-Hα(1)	0.56	GlcN A	C5-H5	0.94
HM2	Cβ-Hβ	0.56	Kdo C	C7-H7	0.96

HM4	Cα-Hα(2)	0.57	Kdo C	C3-H3(2)	0.96
GlcN B	C2-H2	0.63	Kdo D	C6-H6	0.96
GlcN A	C4-H4	0.78	Kdo D	C3-H3(1)	0.97
GlcN B	C3-H3	0.78	Kdo D	C4-H4	1.00
GlcN B	C1-H1	0.79			

Table S2: Chemical shift differences (in ppm) of LPS upon adding PMX-B, PMX-E and PMX-M as derived from the chemical shift mapping experiments. Only changes of chemical shifts, which can be accurately measured, are included.

		Polymyxin B		Polymyxin E		Polymyxin M	
Unit	Atoms	$\Delta^{13}C$	$\Delta^{1}H$	$\Delta^{13}C$	$\Delta^{1}H$	$\Delta^{13}C$	$\Delta^{1}H$
GlcN A	C1-H1	0.084	0.015	0.011	0.018	0.017	0.025
GlcN A	C2-H2	0.015	0.036	0.052	0.021	0.024	0.022
GlcN A	C4-H4	0.312	0.039	0.320	0.090	0.383	0.092
GlcN A	C5-H5	0.154	0.002	0.285	0.024	0.341	0.040
GlcN A	C6-H6	0.109	0.081	0.118	0.084	0.233	0.012
GlcN A	C6-H6	0.038	0.024	0.085	0.004	*	*
GlcN B	C1-H1	0.517	0.031	0.714	0.074	0.329	0.051
GlcN B	C4-H4	0.149	0.129	0.156	0.121	*	*
GlcN B	C5-H5	0.168	0.042	0.186	0.040	*	*
Kdo C	C3-H31	0.552	0.002	0.649	0.009	*	*
Kdo C	C3-H32	0.522	0.037	0.603	0.048	*	*
Kdo C	C4-H4	0.271	0.006	0.287	0.008	0.288	0.006
Kdo C	C5-H5	0.174	0.023	0.182	0.027	0.118	0.034
Kdo C	C6-H6	0.269	0.027	0.280	0.026	0.241	0.076
Kdo D	C3-H31	0.199	0.002	0.179	0.001	0.206	0.002
Kdo D	C3-H32	0.174	0.012	0.182	0.006	0.218	0.014
Kdo D	C4-H4	0.051	0.007	0.042	0.009	0.064	0.006
Kdo D	C5-H5	0.037	0.006	0.023	0.005	0.002	0.001
Kdo D	C6-H6	0.022	0.018	0.021	0.020	0.081	0.018
HM4	Ca-Ha	0.061	0.035	0.196	0.054	0.104	0.000
HM4	Cb-Hb	0.078	0.031	0.083	0.002	*	*
HM4	Cg-Hg	0.013	0.014	0.139	0.000	0.031	0.043
HM3	Ca-Ha	0.000	0.016	0.042	0.022	0.115	0.019

HM3	Cb-Hb	0.078	0.029	0.067	0.025	*	*
HM3	Cg-Hg	0.048	0.007	0.041	0.025	0.021	0.017
HM2	Ca-Ha	0.371	0.007	0.472	0.015	0.468	0.006
HM2	Cb-Hb	0.600	0.120	0.928	0.079	1.068	0.067
HM2	Cg-Hg	0.156	0.005	0.080	0.051	0.285	0.027
HM1	Ca-Ha	0.068	0.029	0.001	0.047	0.105	0.052
HM1	Cg-Hg	0.026	0.001	0.264	0.006	0.353	0.006

Fig. S3: Selected spectral regions of the constant-time [^{13}C,^{1}H]-HSQC of 0.35mM LPS in 300 mM DPC, 40mM acetate pH=4.4, T=310K. Therein, original peak positions are shown in yellow, marked with cross and labeled for the carbon atom. The LPS signals that are shifted upon addition of equimolar amounts of PMX-B, -E and -M are shown in blue, magenta and green, respectively.

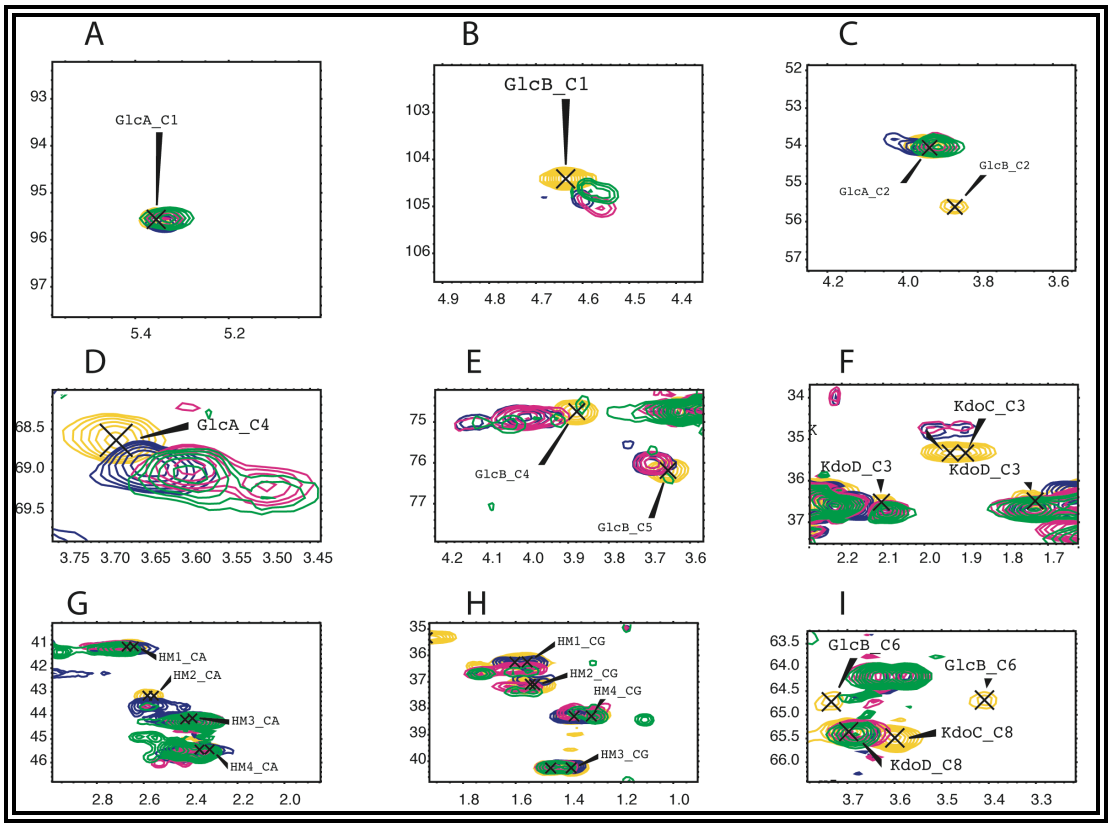
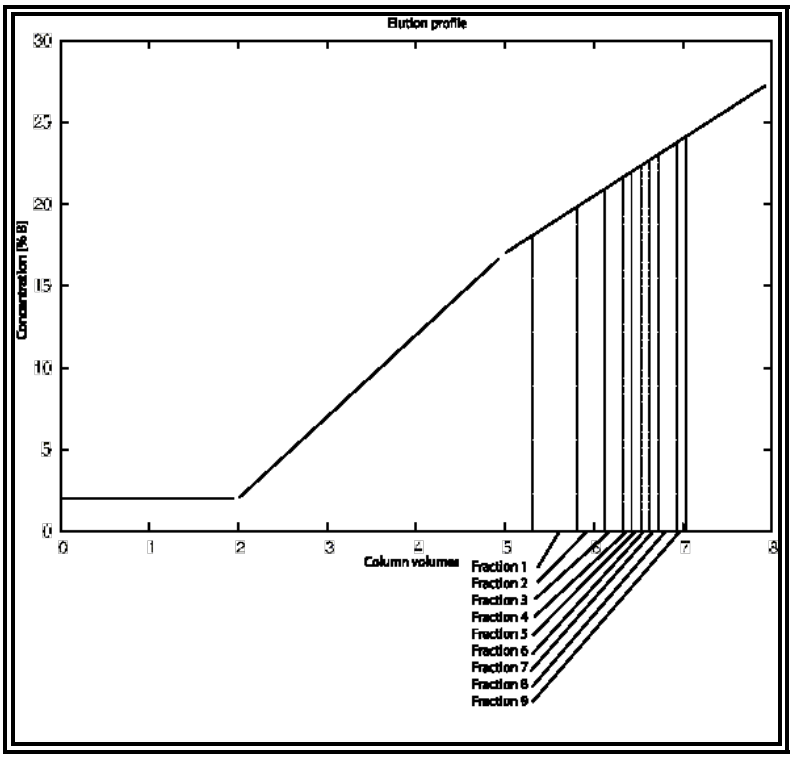


Fig. S4: Elution profile for the HPLC purification of LPS in the ternary solvent mixture showing MALDI MS spectra of selected fractions. The dominant compound of fraction 3 is the form of LPS, which has one phosphate group substituted by pyrophosphate, theoretical MW: 2429.665 for 100% ^{13}C ^{15}N. Fractions 5, 6 and 7 contain the desired form of LPS, theoretical MW: 2349.699.



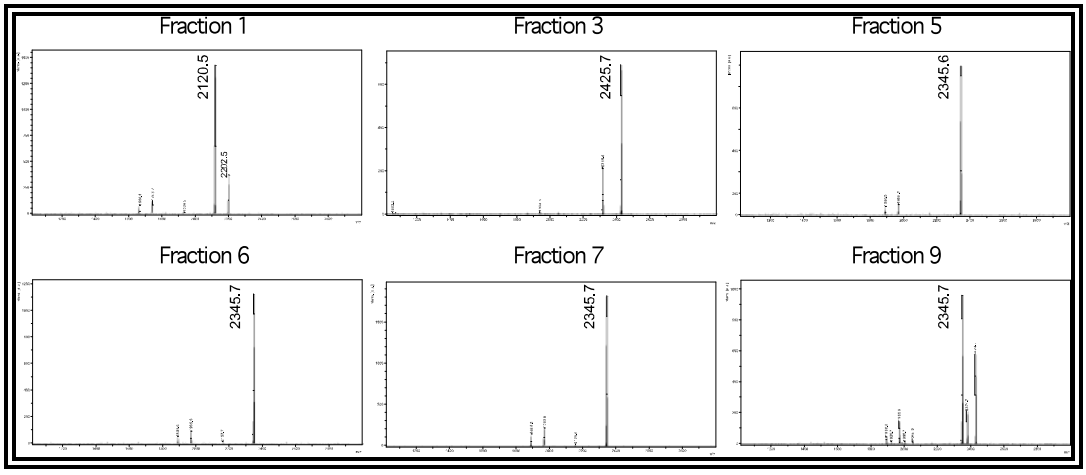


Fig. S5: Anomeric region of the constant-time [^{13}C,^{1}H]-HSQC of LPS. The spectrum is shown close to the noise level to document the purity of the obtained LPS.

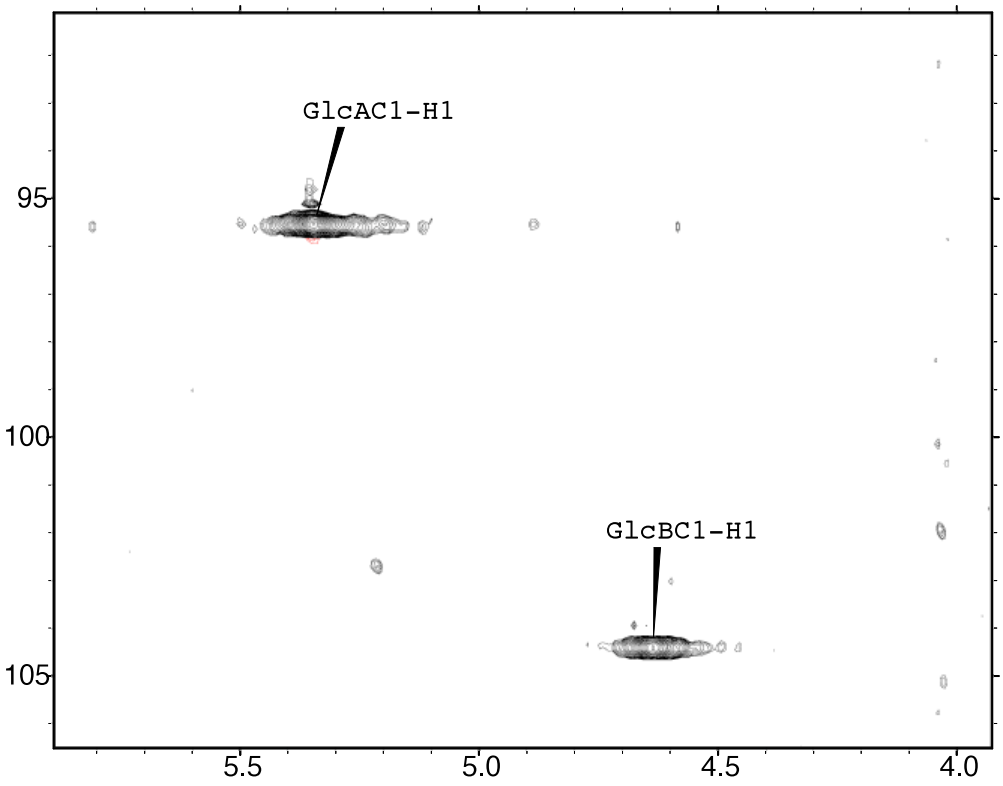


Fig. S6: Differences in LPS chemical shift *changes* when adding PMX between PMX-B vs. PMX-E (left) and PMX-B vs. PMX-M (right).

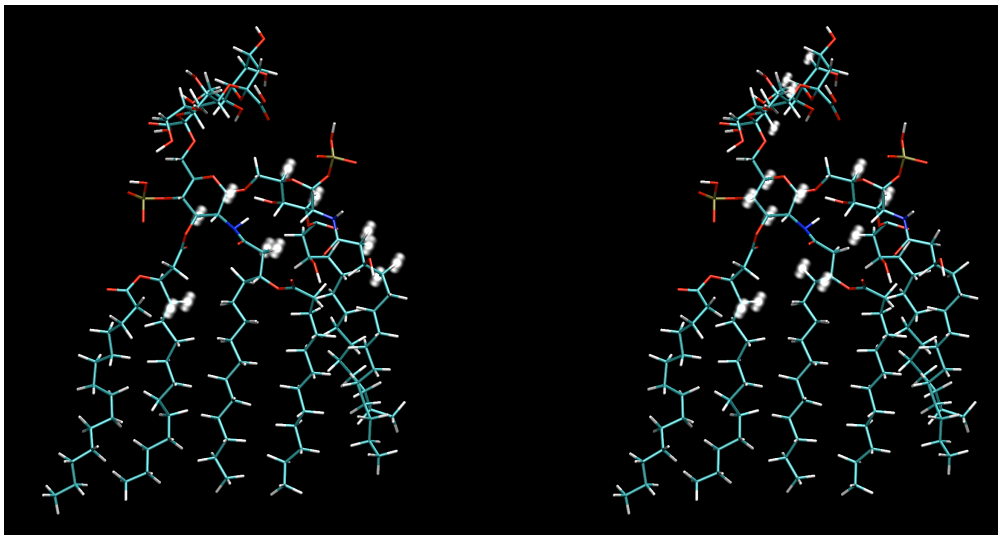


Fig. S7: Structure of PMX-B in the complex with LPS as derived from the MD calculation. Color-coding is according to Figure 6 from the manuscript. Different orientations are shown.

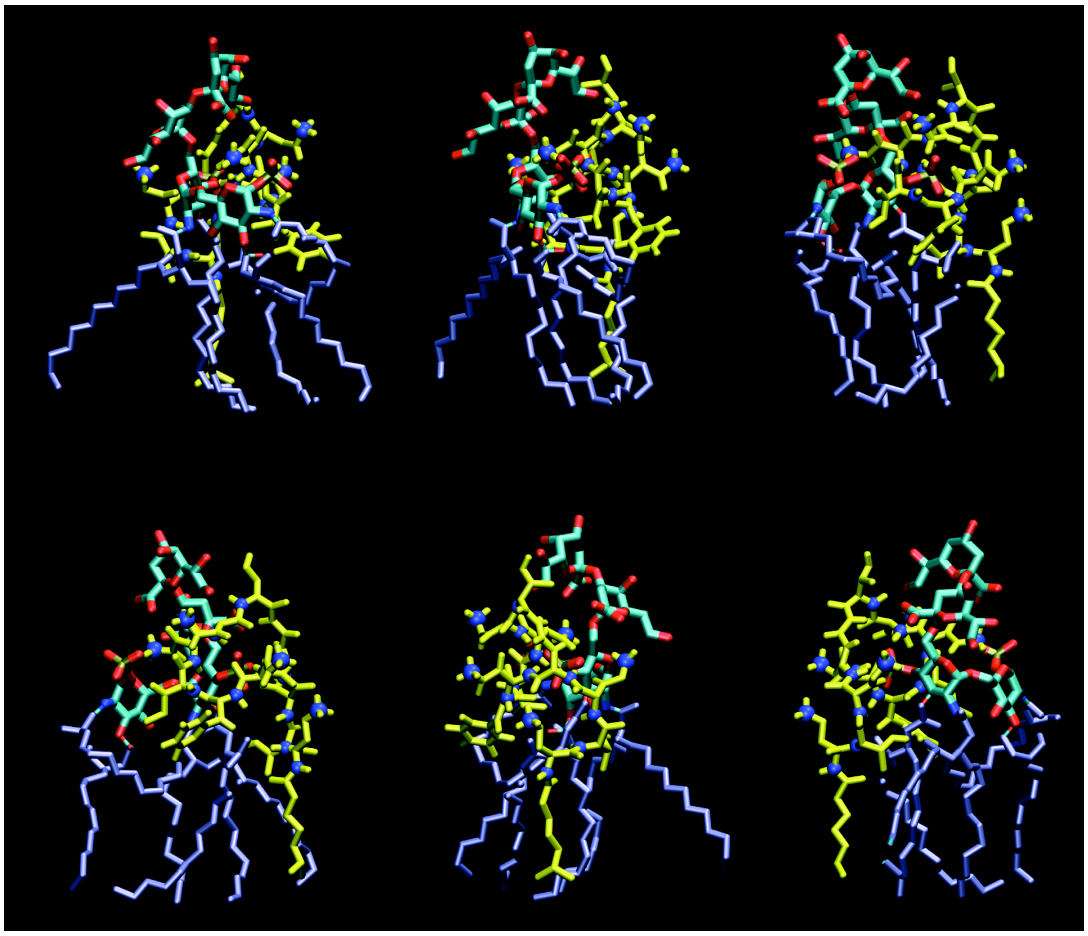


Table S8: Chemical shifts of polymyxin E, 310K, 2mM PMX-E, 300 mM DPC, pH=4.4, 40mM acetate buffer. All chemical shifts were referenced relative to DSS at 310K (FA: fatty acid chain)

	H^N	H^{α}	H^{β}	Others
FA		2.30	1.56	γCH_2 1.29, δCH_2 1.18; εCH_2 1.10;CH_3 0.85
Dab 1	8.49	4.52	2.07, 2.20	γCH_2 3.05, 3.05; δNH_3^+ -
Thr 2	8.23	4.36	4.30-	γCH_3 1.17; γOH -
Dab 3	8.81	4.36	2.14, 2.24	γCH_2 3.08, 3.08; δNH_3^+
Dab 4	8.38	4.15	1.93, 1.99	γCH_2 3.08, 3.45; δNH 7.62
Dab 5	8.24	4.47	2.04, 2.16	γCH_2 3.00, 3.00; δNH_3^+
Leu 6	8.60	4.25	1.52, 1.63	γCH 1.67; Hδ 0.93, 0.91
Leu 7	8.78	4.31	1.60, 1.67	γCH 1.60; Hδ 0.85, 0.93
Dab 8	7.93	4.40	2.14, 2.28	γCH_2 3.12, 3.12; δNH_3^+ -
Dab 9	8.23	4.67	2.21, 2.28	γCH_2 3.05, 3.05; δNH_3^+ -
Thr 10	7.81	4.11	4.28	γCH_3 1.19; γOH -

Table S9: Chemical shifts of polymyxin B, 310K, 2mM of PMX-B, 300 mM DPC, pH 4.4, 40 mM acetate buffer

	H^N	H^{α}	H^{β}	others
FA	-	2.30	1.56	γCH_2 1.30; δCH_2 1.19; CH_3 0.91
Dab 1	8.51	4.53	2.09, 2.20	γCH_2 3.05; δNH_3^+ -
Thr 2	8.24	4.35	4.30	γCH_3 1.19; γOH -
Dab 3	8.81	4.38	2.14, 2.24	γCH_2 3.09, 3.09; δNH_3^+ -
Dab 4	8.38	4.13	1.93, 1.93	γCH_2 3.06, 3.45; δNH 7.62
Dab 5	8.20	4.44	1.96, 2.09	γCH_2 2.90, 2.90; δNH_3^+ -
Phe 6	8.68	4.55	2.98, 3.08	δH 7.26, 7.26; εH 7.16, 7.16; ζH -
Leu 7	8.61	4.19	1.35, 1.49	γH 0.88; δCH_3 0.68, 0.75
Dab 8	7.96	4.39	2.14, 2.25	γCH_2 3.10, 3.10; δNH_3^+ -
Dab 9	8.14	4.39	1.91, 2.01	γCH_2 2.85, 2.85; δNH_3^+ -
Thr 10	7.83	4.11	4.26	γCH_3 1.18; γOH -

Table S10: Chemical shifts of polymyxin M, 310K, 2mM PMX-M, 300 mM DPC, pH=4.4, 40mM acetate buffer

	H^N	H'^{α}	H^{β}	others
FA		2.33	1.59	γCH_2 1.33; δCH_2 1.24; εCH_2 1.34; ζCH_2 1.25; QD 0.87; QE 0.87
Dab 1	8.45	4.56	2.22, 2.12	γCH_2 3.10, 3.10; δNH_3^+ -
Thr 2	8.37	4.25	4.23	γCH_3 1.25; γOH -
Dab 3	8.80	4.45	2.27, 1.71	γCH_2 3.12, 3.12; δNH_3^+ -
Dab 4	8.15	4.27	2.00, 2.00	γCH_2 3.42, 3.16; δNH 7.8
Dab 5	8.34	4.59	2.23, 2.09	γCH_2 3.08, 3.08; δNH_3^+ -
Leu 6	8.77	4.45	1.69, 1.64	γH 1.69; δCH_3 0.99, 0.93
Thr 7	8.56	4.37	4.44	γCH_3 1.19; γOH -
Dab 8	8.35	4.36	2.31, 2.27	γCH_2 3.19, 3.19; δNH_3^+ -
Dab 9	8.79	4.29	2.29, 2.17	γCH_2 3.11, 3.11; δNH_3^+ -
Thr 10	7.94	4.15	4.26	γCH_3 1.22; γOH -

Table S11: Relative attenuations of signals from PMX-B, -E or –M after addition of 5 equiv. of Methyl-5-doxylstearate. The experiments used 4mM unlabelled PMX-B, PMX-E and 0.2mM ^{13}C,^{15}N-labelled PMX-M in 300mM DPC in Acetate buffer pH 4.4, 310K.

Residue Number	PMX-B	PMX-E	PMX-M
1	0.11	0.17	0.03
2	0.71	0.77	0.50
3	0.84	0.80	0.73
4	0.90	0.73	0.75
4(Gamma)	0.50	0.77	0.47
5	0.74	0.84	0.56
6	0	0	0.02
7	0	0	0.22
8	0.57	0.92	0.93
10	0.81	0.69	0.68

NMR sample preparation protocol for the shift mapping experiments

Several milligrams of pure lyophilized LPS powder were accurately weighted, dissolved in deionized water. A small amount of deuterated DPC was added, which allowed to prepare higher concentrations of LPS. This stock solution was aliquoted and lyophilized for further use. (The additional amount of DPC introduced this way was negligible with respect to the amount of DPC in the final NMR sample). Several milligrams of pure lyophilized polymyxins were weighted and stock solutions were prepared by dissolving the peptide in the acetate buffer, which was used for measurement of NMR spectra.

For NMR measurements, an aliquoted amount of LPS was dissolved in the corresponding DPC/acetate solution. Good homogeneity of the sample was achieved by brief (10 s) sonication. In the interaction studies, the amount of acetate buffer, in which DPC was dissolved, was reduced by the volume of added polymyxin solution.In separate experiments we checked that LPS and polymyxins in concentrations used for the sample preparation did not change the pH of the sample. The pH of the sample was not adjusted after the mixing to avoid possible contaminations from the electrode electrolyte components.

Protocol and details for the MD simulation

Setup

In the first step, a set of structures was generated using a simulated annealing protocol. All of the experimentally observed NOEs were incorporated as distance restraints in this calculation. Simulated annealing *in vacuo* allowed fast generation of a diverse set of starting structures that fulfill the experimental constraints. The electrostatic interactions were scaled by a dielectric constant of 20, which roughly mimicks the interface of water/membrane environment.

The set of annealed structures was clustered, so that the most abundant conformations can be selected. A representative conformer from each of the resulting 10 clusters was transferred to the simulation box that included a DPC micelle for further

simulation. LPS-polymyxin complexes were placed into vicinity of the equilibrated DPC micelle consisting of 54 DPC molecules. The subsequent solvation used the same number of water molecules in each case. An artificial atom was placed in the center of the simulation box, which was position-restrained by a strong force. Distance restraints were between this atom and the atoms at the center of the aliphatic chain of each DPC molecule as well of the LPS lipid chains were applied.

The restrained distance corresponds to the approximate average distance of these atoms from the center of the micelle. The distance restraints were time averaged giving the system sufficient motional freedom. The experimental distance restraints between LPS and polymyxin were strongly kept to preserve the structure of LPS-polymyxin complex. The system was driven close to equilibrium within half a nanosecond MD simulation. Each system was then placed into a smaller simulation box and solvated by an accordingly smaller number of water molecules. The distance restraints were then released and the system was allowed to equilibrate within a 2 ns simulation, after which a short production MD simulation followed.

Improvement

To further improve agreement with experimental data a final structure from the trajectory described above that gave the best score was used to create new set of diverse structures. Therein, simulated annealing cycles were calculated, in which the DPC molecules were very weakly position restrained while atoms of LPS and polymyxin were allowed to move freely except for the terminal atoms of aliphatic chains of LPS. The water was removed and replaced by a implicit solvent model. This setup generated a set of diverse conformations, in which steric interactions of atoms of LPS or PMX-B with DPC as well as electrostatic interactions were reasonably taken into account. At the end of this procedure a set of systems was established that included the DPC micelle into which the LPS-polymyxin complex was embedded. During this step, distance restrains were exclusively applied to specific interactions involving those aliphatic LPS atoms for which significant changes were observed by chemical shift mapping. Since the interaction partner atom of PMX-B was unknown, various possibilities were tested. In all

cases, long-range distance restraints of Phe6 ring atoms to atoms of the fatty acids were used. From a set of 450 annealed structures 45 were selected according to the occurrence of experimentally verified interactions. These were then further subjected to a short (120 ps) equilibration simulation with explicit solvent. Five trajectories were finally computed without restraints and a representative snapshot from the most stable trajectory is shown in Figure 6 in the main text.

Analysis

Each of the resulting MD trajectories was assessed in several steps. First, RMS deviations with respect to the start of the production trajectory were calculated to monitor, whether the MD trajectory is stable. All atoms of LPS and polymyxin were used for RMSD calculation. These trajectories were derived from the conformations based on NOE contacts, therefore another experimental information was required to evaluate, how they reflect the experimental system. Average distances between each atom of LPS with each atom of the polymyxin were calculated over the MD trajectory. Those distances were selected for which participating atoms displayed changes in the chemical shift mapping experiments. All the analysis was done using tools from the GROMACS MD package and by own scripts. Structures were visually inspected by VMD, which was also used to prepare figures.

Simulation parameters

Vacuo simulated annealing

All the interactions were computed without any cutoff for reasons of computational stability, the dielectric constant was set to 20, the integration step to 0.5 fs, the force constant for distance restraints to 1000 kJ mol-1 nm-2. The annealing cycle had the following profile: 80 ps at 1200K, linear cooling to 298 over 5 ps, 55 ps at 298 K, linear cooling to 0 K over 50 ps and the cycle was closed by linear heating to 1200K over a time of 10ps.

Simulated annealing with implicit solvent

The atoms of DPC molecules were position restrained with a force constant of 20, terminal atoms of LPS aliphatic chains with force constant 100. The dielectric constant was set to 80 and the GBSA implicit solvent model as implemented in GROMACS 4.0 was used. The simulated annealing cycle was shortened 4 times to save computer time.

Equilibration and production MD

For initial equilibration the system of DPC, LPS and polymyxin was solvated with 12574 SPC water molecules. Equilibration after simulated annealing with implicit solvent and the production MD was performed with 7295 water molecules. One Cl^- ion was added to neutralize the net charge. In the initial equilibration and production MD, the integration step was set to 1 and 2 fs, respectively. The simulations were performed in the NPT ensemble using periodic boundary conditions. Long-range electrostatics were treated by using Particle-mesh Ewald summation. The temperature was weakly coupled (coupling time 0.1 ps) to T=310K. Bond lengths were constrained using the LINCS algorithm. The pressure was weakly coupled (coupling time 0.5 ps, compressibility 4.5e-5 bar) to 1 bar. The force constant of distance restraints during the equilibration was set to 5000.

Chapter 5

Studies of the structure of the N-terminal domain from the Y4 receptor, a G-protein coupled receptor and its interaction with hormones from the NPY family

5.0 Introduction

Neuropeptide Y receptors, also called as Y receptors, are members of the rhodopsin-like G-protein coupled receptor (GPCR) family 1b. Neuropeptide Y receptors occur in multiple forms. At least five Y receptors have been cloned from mammals (Y1, Y2, Y4, Y5, and y6). These PP-fold receptors display many of the structural features of rhodopsin-like receptors. It is believed to form a disulfide bond between extracellular loop (E- I and E- II). In case of Y1, Y4, and y6 they also have two additional cysteines, one in the amino-terminal tail and one in E-III that potentially can form a second disulfide bound. They have one or several consensus sites for N-linked glycosylation (Asn-X-Ser/Thr). The Y1 and Y4 receptors have three such glycosylation sites located on the extracellular side in the amino terminus and one in the second extracellular loop whereas Y2 has only one and Y5 receptors have two glycosylation sites in the amino terminus, respectively. All these receptors have at least one cysteine in the cytoplasmic tail that probably anchors the tail to the inside of the membrane by palmitoylation [1]. The amino acids which form the α - helix after the 7 th transmembrane region in bovine rhodopsin are well conserved among the PP-fold receptors suggesting that this structure may be present in these Y- receptors too.

The Y-receptors are activated by neuropeptide Y, pancreatic polypeptide, and peptide YY. These three peptides belong to one family of neuroendocrine hormones, the so-called NPY hormone family. They are 36-residue peptide amides and are rich in tyrosine (Tyr) residues. NPY is widely distributed throughout the mammalian brain and is one of the most potent orexigenic factors[2]. PYY exists in two major forms: PYY (1–36)

and PYY(3–36). PYY(3–36) is a peripherally active anorectic signal. PP is a hormone that is released primarily from the pancreas in response to food intake. NPY analogues, as well as PYY and PP, exhibit varying degrees of affinity and specificity for these Y receptors. NPY and PYY posses a similar pharmacology displaying nanomolar affinities for all receptor subtypes[3], whereas PP binds with very high affinity and selectivity to the Y4 receptor and hence it also called as PP- preferring receptor[4]. The Y4 receptor consists of 375 amino acids and it has very low degree of sequence identity between species and is found to be one of the most rapidly evolving GPCRs known[5, 6]. Human Y4 mRNA is mainly expressed in the colon, small intestine and prostate, although other peripheral tissues appear to lack it, and various CNS regions display low expression levels[5]. The Y4 receptor is involved in the regulation of food intake and motility of the gastrointestinal tract.

In this work we focus on structural studies of the isolated 41 residue N terminus of the Y4 receptor. The location of this segment in the context of the entire human Y4 receptor is shown in snake plot in Fig. S1 (see Supp. Mat.). In addition, we investigate possible interactions with the PP fold hormones both qualitatively and quantitatively. By limiting the system of the study to just the N-terminal domain and with the help of various biophysical methods we were able to develop a rather detailed picture, that would presently be difficult to achieve using the entire receptor. Moreover, we report on the synthesis of the difficult to express N-terminal domain suggesting a generally useful method to produce these polypeptides in isotopically-labelled form. The structure of N-Y4 and its topology in the presence of DPC or SDS micelles was elucidated by high-resolution NMR techniques. While unstructured in solution, in the presence of micelles a hydrophobic segment associates with the micelle and folds into a a-helix. Chemical shift mapping revealed potential interaction sites between PP and N-Y4. SPR techniques quantified the strength of this interaction. Mutagenesis studies identified residues of PP that are likely to be important for binding N-Y4. The data indicate that the isolated N- Y4 is capable of weakly binding to PP, and that much of the binding affinity is due to electrostatic interactions. To simulate the receptor milieu the carboxyl terminus of N-Y4 was additionally conjugated to a C12 fatty amino alcohol (dodecylphospho-

ethanolamine) chain thereby mimicking its conjugation to the first TM helix in the entire receptor. In this lipopeptide the structure of the N-Y4 was not significantly affected. The study shows that PP associates to the flexible, central segment of N-Y4 and we speculate that transient binding to the N-terminal domain may facilitate transferring PP from the membrane-bound state into the receptor-binding pocket.

5.1 Results

5.1.1 Recombinant production of N-Y4

The N terminus of the Y4 receptor comprises 41 residues and is highly water-soluble. However, attempts to express it in form of a soluble ubiquitin fusion in *E. coli* resulted in unspecific fragmentation. To circumvent this problem, the N-Y4 was expressed as a fusion to the highly insoluble protein ketosteroidisomerase (KSI), which resulted in accumulation of the fusion protein in inclusion bodies. A TEV protease cleavage site was introduced to facilitate removal of the fusion partner[7, 8]. The sequence recognized by the TEV protease is ENLYFQ with Q as the P1' residue. To achieve the natural peptide sequence after cleavage, the P1' residue was replaced with the first residue from the target sequence (here it is Met)[9], and an additional GSGSGS linker was inserted to prevent steric hindrance during cleavage.

A problem of the chosen strategy was that the fusion protein must be solubilized in detergent that is compatible with the active protease. After extensive detergent screening, we observed that the ionic detergent sarcosyl solubilizes the fusion protein while preserving TEV protease activity to some extent. As shown in Fig. 1 cleavage efficiency is around 40% allowing recovery of about 2mg of ^{15}N-labeled N-Y4 from 1L of culture.

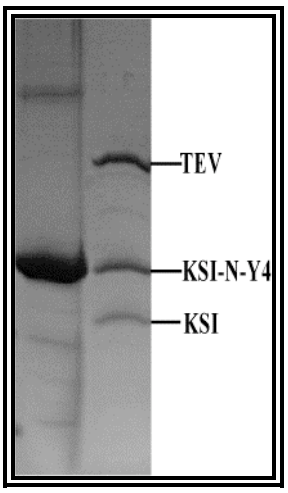


Figure 1: SDS-PAGE of the cleavage product of the ketosteroid isomerase- N-Y4 fusion after cleavage with the TEV protease. A size marker is shown on the left. Note that N-Y4 due to its small size cannot be detected on the gel.

5.1.2 The structure of N-Y4

Although the size of the N-terminal domain is rather small, reduced chemical shift dispersion due to the fact that the peptide in water is largely unstructured complicated its analysis. Nevertheless, using 3D ^{15}N-resolved NOESY and TOCSY spectra it was possible to assign the ^{15}N,^{1}H-correlation map. Furthermore, no NOE crosspeaks between amide protons could be detected. Recording a second set of 2D and 3D spectra in the presence of DPC micelles resulted in large chemical shift changes in some parts of the sequence (see Fig.2). Moreover, sequential NOEs between amide protons as well as Ha,Hb (i,i+3) contacts usually only observed in helices were seen (see Supp. Mat.). A structure calculation using restraints derived from the NOESY spectra revealed the presence of a helical stretch encompassing residues 5 to 10 (shown in Fig.3).

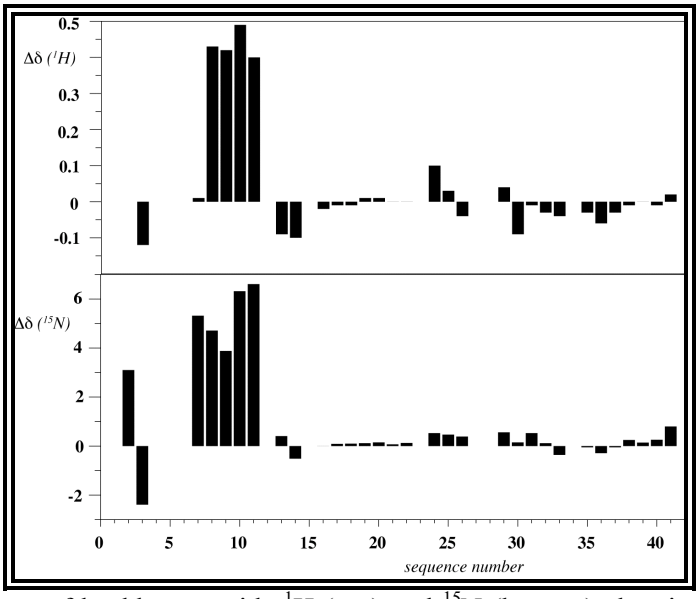


Figure 2: Differences of backbone amide ^{1}H (top) and ^{15}N (bottom) chemical shifts of N-Y4 in the presence and absence of DPC micelles.

To verify formation of stable secondary structure $^{15}N\{^{1}H\}$-NOE spectra were recorded both in the absence as well as in the presence of DPC. The heteronuclear NOE sensibly reports on the rigidity of the backbone at the corresponding residue, with negative values characteristic of flexible parts and values larger than 0.5 usually observed in elements of secondary structure. The $^{15}N\{^{1}H\}$-NOE data show dramatic differences in aqueous medium and DPC. Residues 1-27 have values <0 for N-Y4 in water whereas all of these residues have $^{15}N\{^{1}H\}$-NOE values >0 in the DPC bound state (Fig. 4). Strikingly, residues 5-10 have a $^{15}N\{^{1}H\}$-NOE >0.5. Interestingly, a segment encompassing residues 26 to 33 is rather rigid, in both environments. We observed sequential amide proton contacts in that region for almost all residues, but the corresponding Ha, Hb (i,i+3) contacts were generally missing. When comparing chemical shifts of amide protons in the two environments the largest differences were observed in that segment that obviously becomes structured in the presence of the micelle, indicating the presence of a nascent helix in that part. To conclude, the N terminus is largely unstructured in the absence of a membrane whereas a short helical stretch comprising a hydrophobic segment in the N terminus of the sequence is formed in presence of DPC micelles.

Figure 3: Comparison of the structures calculated for N-Y4 in the presence of DPC (left) or SDS (right) micelles (only bonds from backbone atoms are depicted). Bonds from disordered residues 16-41 are not shown for clarity.

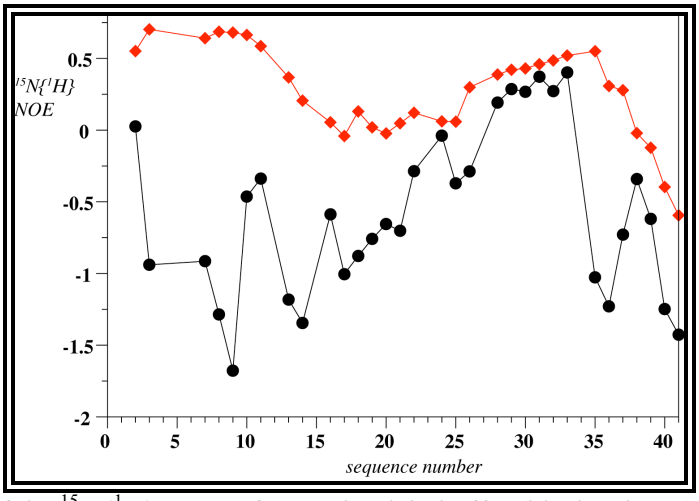


Figure 4: Values of the $^{15}N\{^1H\}$-NOE of N-Y4 in plain buffer (black spheres) and in the presence of DPC micelles (red diamonds). Data were recorded on 1mM samples at pH=5.6, 310K, at 700 MHz proton frequency.

Considering the importance of electrostatic interactions for ligand binding and to investigate whether (stabilizing) interactions of the N-terminal domain with the membrane head groups might be formed we further initiated structural studies of N-Y4 in the presence of SDS micelles, a negatively charged membrane mimetic. Values of the $^{15}N\{^1H\}$-NOE rapidly revealed that N-Y4 was not significantly better structured in this environment. Moreover, a structure calculation again revealed the presence of an a-helix spanning the region between residues 3 to 10. NOEs between sequential amide protons were seen at the C-terminal end from residue 36 on, but the corresponding Ha,Hb (i,i+3) contacts were missing, indicating that a transient helix is formed towards the C terminus.

Interestingly, this part in the full-length receptor is connected to the first TM. In general, sequential amide proton contacts in the more flexible regions were stronger when compared to the data recorded in the presence of DPC suggesting that the negatively charged surface promotes the formation of transient helical structures to a slightly larger extent. This fact is particularly well-documented in the heteronuclear NOEs for residues of the segment encompassing residues 19-25, which is much less flexible in the presence of SDS micelles (see Fig. S8). But in general the structural features of the peptide in DPC and SDS were similar (for more data on the SDS-recorded sample see the Supp. Mat.)

5.1.3 Topology of membrane-association

The proximity of protons of the N-terminal domain to the micelle surface was probed by using micelle-integrating spin labels. The paramagnetic moiety of 5-doxyl stearic acid was shown to reside in the head group region[10]. Consistent with the assumption that structuring of the N-terminal segment is induced by binding to the micelle, signals from the amide moieties within that segment experienced the largest signal reduction (see Fig. S9). The spin-label data indicate that the N-terminal helix is tightly associated with the micelle, whereas the central segment makes more transient contacts. Motions in that region are likely limited at both ends by the adjacent hydrophobic residues 24-30 and the membrane-anchored N-terminal helix. It was previously demonstrated that attenuations in helical regions of surface-associated peptides follow periodic patterns[11, 12]. The present data indicate that the helical region is not bound in a parallel fashion to the micelle-surface. Moreover, from the lack of a clear pattern in the attenuation we conclude that this part is also not anchored in a precisely defined mode.

We have additionally tested whether binding of bPP to N-Y4 could possibly trigger dissociation of the N terminus from the micelle. However, no decrease of signal reduction from the spinlabel could be detected upon addition of bPP to micelle-bound N-Y4, indicating that N-Y4-micelle contacts are largely unchanged, even in the presence of a large excess of bPP (concentration ratio of N-Y4 to bPP 1:30) (data not shown). This

indicates that bPP cannot initiate detachment of N-Y4 from the micelle surface, supporting the view that the contact site between bPP and Y-4 is not located in the helical segment of Y-4 and hence does not interfere with micelle association.

5.1.4 Immobilizing the N terminus on the membrane

In the native Y4 receptor the segment that has been studied in this work is connected to the first TM helix. In order to address whether anchoring of N-Y4 at its C-terminal end to the membrane influences the structure or the binding properties of the N-terminal domain a lipopeptide was chemically synthesized, in which receptor residues 1-41 were covalently linked at their C-terminus to dodecylethanolamine to provide stable anchoring of the lipopeptide in the micelles. The lipopeptide was prepared using standard amino-acid coupling chemistry, purified, and could be tightly integrated into the DPC micelles. A superposition of the NOESY spectra of N-Y4 and the lipopeptide in the presence of DPC micelles revealed that chemical shift differences are exclusively observed in vicinity of the lipid attachment site. Moreover, cross peaks between amide protons occur at identical positions, indicating that the secondary structure of both the peptides is highly similar. To conclude, anchoring of N-Y4 onto the micelle does not influence its secondary structure, which more likely is determined by partitioning of residues of the hydrophobic Leu-rich segment into the membrane. As evident from Fig. 3 the carboxyl terminal segment of N-Y4 possesses high flexibility both in the presence and in the absence of DPC micelles. Whether this will also be true when the C-terminus is linked to the first TM helix is presently under investigation.

5.1.5 Interaction between N-Y4 and neuropeptides from the NPY family

Possible interactions between peptides from the NPY family and N-Y4 were probed both by chemical shift mapping as well as by surface plasmon resonance (SPR). PP represents a natural ligand for the Y4 receptor, and accordingly the binding affinity between N-Y4 and PP was measured under physiological conditions (10mM HEPES pH 7.4, 150mM NaCl) both in absence and presence of DPC micelles. The data for chemical shift mapping were acquired using ^{15}N-labeled NPY, PP or PYY and unlabeled N-Y4 as well as using ^{15}N-labeled NY-4 and unlabelled neuropeptides. The shift mapping

experiments revealed significant shift changes in the PP-N-Y4 interaction studies (see Fig. 5). Large changes in the PP/N-Y4 system occurred close to positions that were later on shown to be sensitive to replacement by Ala residues (vide infra). In addition, the shift changes involving PYY and NPY are generally much smaller compared to those with PP (data not shown).

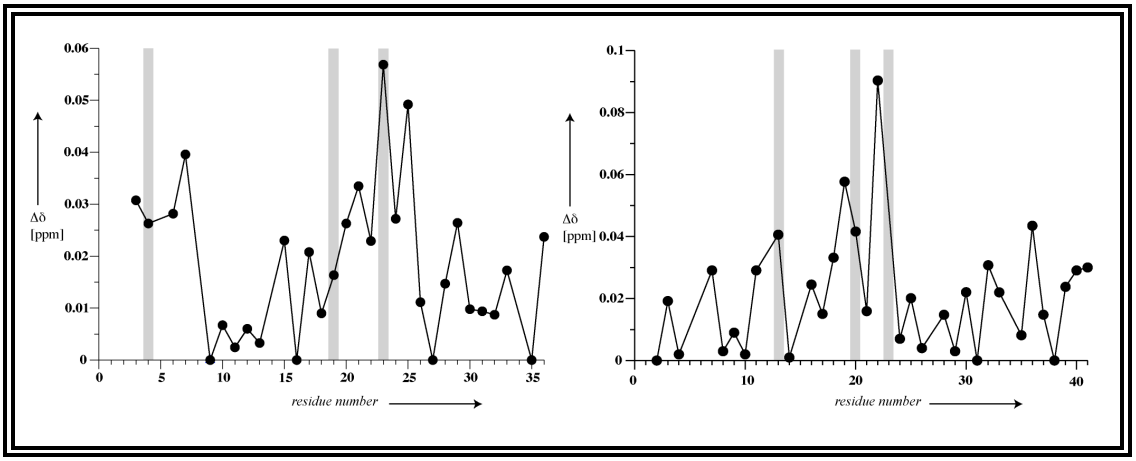


Figure 5: Differences of chemical shifts of amide proton and nitrogen frequencies of backbone resonances of bPP in the presence and absence of N-Y4 (Δδ= δ(bPP(N-Y4)) – δ (bPP) (left) and of N-Y4 upon addition of bPP (right). Values are computed according to $\Delta\delta_c(^1H,^{15}N) = \text{sqrt}\ [(\Delta\delta^1H)^2 + 0.2X(\Delta\delta^{15}N)^2]$. Positions at which mutations were performed (E4K, Q19R and E23A in PP and K13A, R20A and K23A in N-Y4) are indicated by grey bars.

The strength of the interaction of PP with N-Y4 was quantified by SPR in absence of detergent. Therein, the N-terminally biotinylated neuropeptides were immobilized on a Streptavidin-coated chip, and the cells were flushed with solutions of N-Y4 (see Fig. 6). The K_D derived from both kinetic and steady-state analysis was 50 μM for bPP, whereas binding affinity for NPY and PYY was too low to be measured with this technique (> 1mM).

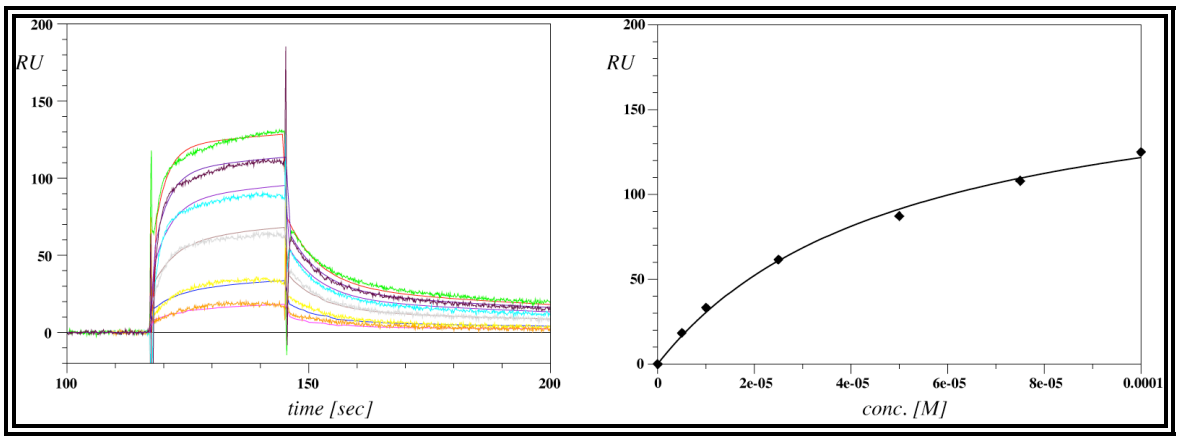


Figure 6: Left: SPR sensogram of the interaction of N-Y4 with bPP for various concentrations of N-Y4 (in the range of 5 to 100 μM). Right: Plot of the steady-state value of the sensograms *vs.* the concentration of N-Y4, used for extraction of the dissociation constant K_D.

Measuring binding of membrane-immobilized peptides towards N-Y4 by SPR methods is technically very challenging, and hence K_D in the presence of micelles were measured using NMR data by fitting changes in chemical shifts as derived from peak positions of the neuropeptides in [^{15}N,^{1}H]-HSQC spectra in the presence of varying amounts of N-Y4. For micelle-bound bPP the K_D to N-Y4 is approx. 600μM and experiments, in which varying amounts of PP were added to N-Y4, resulted in a very similar value. Apparently, the K_D in the presence of micelles is much lower than in the absence of micelles. This is not really surprising since it presents the affinity of the ligand towards the N-terminal domain in the presence of competing membrane binding, and hence reflects the difference in binding affinity between the two sites.

NPY and PYY possess 80% sequence identity between each other[13], while PP only shares about 50% homology to each of them. All these neuropeptides display a remarkable separation of charges along the sequence: The positively charged residues occur in the C-terminal half of PP from almost all organisms sequenced so far (see Table 1). In order to identify residues that may contribute significantly to the different pharmacological profiles of NPY/PYY and PP at the Y4 receptor we have aligned the sequences. Particular attention was paid to charged or aromatic residues that are known to be generally involved in GPCR-ligand interactions. The N termini of all Y receptor subtypes are generally negatively charged with the exception of N-Y4 that contains a net

positive charge (see Table 1). Considering the high number of positive charges in N-Y4 and negative charges in the N-terminal half of bPP electrostatic interactions are likely to be responsible for binding, and such forces are also expected to result in the observed rather weak binding affinities.

Table 1. Sequence alignment of the principal members of the NPY family and of the N-terminal domains from the various Y receptor subtypes. Positions in bPP and hN-Y4 replaced by other amino acids in this work have been underlined.

pNPY: YPS<u>K</u>PDNPGE DAPAEDMA<u>R</u>Y YS<u>A</u>LRHYINL ITRQRY-NH_2
pPYY: YPA<u>K</u>PEAPGE DASPEELS<u>R</u>Y YA<u>S</u>LRHYLNL VTRQRY-NH_2
bPP : APL<u>E</u>PEYPGD NATPEQMA<u>Q</u>Y AA<u>E</u>LRRYINM LTRPRY-NH_2
* * * * * * * ** ** **
hN-Y1: MNSTLFSQVE NHSVHSNFSE KNAQLLAFEN DDCHLPLAMI
hN-Y2: MGPIGAEADE NQTVEEMKVE QYGPQTTPRG ELVPDPEPEL IDSTKLIEVQ
hN-Y4: MNTSHLLALL LP<u>K</u>SPQGEN<u>R</u> S<u>K</u>PLGTPYNF SEHCQDSVDV M
hN-Y5: MSFYSKQDYN MDLELDEYYN KTLATENNTA ATRNSDFPVW DDYKSSVDDLQ

As depicted in Table 1 common acidic residues in PP, NPY and PYY are located at positions 6, 10 and 15. PP mutants E4K, Q19R and E23A were produced by site-directed mutagenesis in order to probe for the importance of differently charged residues between PP and NPY/PYY at these positions. The dissociation constant for Q19R-bPP was only marginally reduced to 89 μM, whereas binding of E4K-bPP and E23A-bPP to N-Y4 was too weak to be detected by SPR. The data indicate that it is the additional negative charges in PP and their distribution along the sequence that may be important for its different binding affinities at the N-Y4.

In order to verify that electrostatic interactions between acidic residues of PP and basic residues in the N-Y4 are contributing to binding, the K13A, R20A and K22A mutants of the N-terminal domain of the Y4 receptor were synthesized and investigated by SPR. In all of these mutants binding to bPP was significantly reduced. The measured values for the K_D were 249 μM (R20A), 281μM (K22A) and for K13A binding was too

weak to be detected by SPR. The combination of the mutagenesis studies performed on acidic residues of PP and basic residues of N-Y4 suggests that the binding affinity between the two is determined by electrostatic interactions to a large extent. In this work we have abstained from experiments in which residues in PP and N-Y4 were charged-reversed simultaneously because in those mutants electrostatics are likely to be perturbed in both molecules, and hence it is questionable whether activity could have been rescued.

5.2 Discussion

The mechanism for recognition of ligands by their receptors is of prime biological and pharmaceutical interest. Due to the enormous problems in expression, purification and reconstitution of sufficient amounts of GPCRs, little progress has been made in structural studies over the last decade, and so far bovine rhodopsin and the b-adrenergic receptor are the only GPCRs for which high-resolution X-ray data are published. In this work we have attempted to investigate the structure of the isolated N-terminal extracellular domain of the Y4 receptor, a GPCR targeted by hormones of the NPY family, and which binds to PP with very high affinity. Moreover, we determined the interaction with PP and the other members of the NPY family and investigated the role of specific residues for binding.

Structural studies of GPCR fragments could possibly suffer from the fact that interactions with the remainder of the receptor are missing that may be structurally relevant. As to the present analysis the N-terminal domain of the published crystal structure of b-adrenergic receptor was largely unstructured, and did not display interactions with other parts of this GPCR, in particular not with the extracellular loops. This supports our contention that the conformations of the N-terminal domains of a GPCR are not significantly determined by interactions with the remainder of the receptor. Such a study also allows us to directly define contributions of residues from the N-terminus of the Y-4 receptor to ligand binding. While the N-terminal domain of Y4 is largely unfolded in solution upon binding to zwitterionic (DPC) or negatively charged (SDS) micelles, a hydrophobic segment comprising residues 5 to 10 forms a rather stable a-helix, and the nascent helix encompassing residues 26-35 is slightly rigidified.

The central region and the C-terminal hexapeptide remain largely unstructured. The helical segment comprising residues 5 to 10 is entirely formed by hydrophobic residues. The structural data and the internal backbone dynamics of N-Y4 in the presence of zwitterionic (DPC) and anionic (SDS) headgroups display only minor differences indicating that the conformation does not depend on specific features of the surrounding lipids. Both formation of secondary structure and association with the membrane seem to be controlled by the hydrophobicity of the residues and their partitioning into the membrane[14]. Strongly favorable values for the latter are encountered only in the a-helical stretch and in the segment between residues 24 to 30, exactly those regions for which the spin-label data indicate proximity to the water-membrane interface. Spin-label, dynamics and structural data of Y-4 reveal the central segment to be rather flexible. The segregation of N-Y4 into structured and flexible regions is very similar in the presence of zitterionic or negatively charged lipid headgroups. As a consequence of these features it appears likely that this domain may perform larger movements on the membrane surface, and hence could possibly undergo various structural or translational transitions in order to interact with the extracellular loops or with the membrane-bound ligands. We like to mention at this point that the N-terminal domain of the b-adrenergic receptor was also disordered in the crystal structure from Kobilka[15, 16], and that the N-terminal domains from many other class-1 GPCRs are predicted to be largely unfolded. This indicates that the fact that N-Y4 is mainly flexible is likely not an artifact due to the usage of a receptor fragment but rather reflects a commonly encountered feature of these receptors.

We have recently proposed that binding of hormones from the NPY family to their receptors is preceded by association of the ligands to the membrane. According to ideas originally proposed by Kezdy and Kaiser[17, 18] and later developed into the membrane-compartment model by Schwyzer[19, 20] binding to the membrane reduces the search for the receptor to two dimensions, increases the concentration in the vicinity of the receptor and possibly induces conformations that facilitate receptor binding. Structural studies of porcine (p) NPY[11] and PYY[21] and of bovine (b) PP[22] bound to membrane-mimicking dodecylphosphocholine (DPC) micelles revealed large structural changes occurring during membrane association[23]. From this picture the

important question arises how the hormones enter the binding pocket, once the membrane-bound species has laterally diffused along the membrane into the proximity of the receptor. The seven-helix bundle provides a rather rigid scaffold that does not allow large rearrangements of the extracellular loops in order to facilitate diffusion of the membrane-bound ligand into the binding pocket. Therefore the hormones need to detach from the membrane. Data for binding affinities of the hormones towards phospholipid membranes determined by us using SPR indicate that membrane binding is only moderate[21]. Any part of the receptor that possesses higher affinity to the peptides than the membrane does, and which could be accessed by a ligand that is in proximity to the membrane surface, may help to guide the ligand into the binding pocket. The N-terminal domains of the Y receptors are 40-50 amino acid residue long polypeptide segments located in the extracellular space[24], and hence present potential interaction sites for the ligands. This work now indicates that at least for PP transient association with the N-Y4 may be part of the cascade of events leading to receptor activation. It should be emphasized here that transient binding to the N-terminal domain does not exclude larger structural changes in the conformations of loop residues that may occur later on when the ligands have diffused into the genuine receptor binding pockets. Such changes or rotations of the TM helices are believed to be important for receptor activation, and the above-described events merely serve to guide the ligand from the membrane-bound state into the binding pocket.

Binding of PP to the N-terminal domain of the Y4 receptor, which is often referred to as the PP-preferring receptor, is moderate with a dissociation constant of about 50μM. NPY and PYY, two hormones from the NPY family with very similar pharmacology and high sequence similarity with respect to each other, do not bind to this domain. Sequence alignments reveal that PP overall is more negatively charged than NPY or PYY, particularly in the N-terminal region, and our studies show that replacement of E4 or E23 in PP largely abolished binding to N-Y4. Furthermore, introduction of Arg into position 19 lead to only marginal changes in binding affinity. The N-Y4 domain, in contrast to the N-terminal domains from all other receptor subtypes, contains a comparably large number of positively charged residues (K13, R20

and K22), which are also relatively close to each other in sequence. Their replacement by Ala as described above leads to significant losses in binding affinity. To conclude taking the importance of acidic PP and basic N-Y4 residues into account we speculate that electrostatic interactions between PP and N-Y4 are crucial for this interaction. However, it must be emphasized that *a priori* it is not clear in our case whether residues from the N terminus are interacting with residues from the extracellular loops thereby modulating the effective charge experienced by the peptides. This question can only be addressed experimentally with confidence when structural studies of the full-length receptor in a functional state become available.

Unfortunately, not much pharmacological data is available for the entire Y4 receptor. In case of the human Y1 receptor an Asp residue at the interface between TM helix 6 and the third extracellular loop was proposed to contribute largely to binding NPY[25] in the full-length Y1 receptor. Considering that Asp at this position is conserved amongst all Y receptor subtypes it was speculated that this residue generally contributes to binding in all subtypes. Nicole et al. investigated the role of this $Asp^{6.59}$ in more detail[26] and verified the proposed interaction of Arg33 or Arg35 with acidic third extracellular loop (ECL3) residues in the other Y receptor subtypes. Our data now indicate that in addition to the above-described interaction additional contacts between acidic residues of PP and basic residues of the N-terminal domain of the Y4 receptor may contribute to binding. Association of the N-Y4 with PP may be therefore not only be of transient nature helping the ligand to be transferred from the membrane-bound state into the receptor binding pocket, but may also exist in the ligand-bound state, contributing to the high binding affinity and selectivity of PP at the Y4 receptor.

5.3 Conclusions

Based on the data described above, we speculate that the N-terminal domain of the Y4 receptor may help in transferring PP from the membrane-bound state into the receptor-binding pocket (see Fig. 7). As proposed by us in case of ligands of the Y receptors[23] PP initially associates with the membrane. By binding to the membrane the effective concentration in vicinity of the receptor is increased, the search is reduced from

three to two dimensions, and conformations closer to those of the bound state may be induced according to the membrane-compartment model[19, 20]. Biacore data of PP binding to phospholipid surfaces indicated that binding to membranes is moderate[27]. Accordingly, an equilibrium is formed, in which PP rapidly diffuses on and off the membrane, but mostly remains in vicinity of the membrane. When PP has diffused into proximity of the receptor where interactions with the latter can occur it may transiently bind to N-Y4 from solution. Whether the complex of PP and N-Y4 itself will move into vicinity of the extracellular loops, or whether the position of N-Y4 is fixed by interactions with the membrane or the remaining portion of the receptor is presently unclear.

A scenario, in which N-Y4-bound PP would be transferred into the binding pocket by a translational movement of parts of the N-terminal domain is at least compatible with the experimental data. These indicate that the binding region for PP is located in its central segment, which at the same time is the only part of N-Y4 that is not making significant contacts with the membrane surface, and which also possesses sufficient internal flexibility to allow the necessary movements. We presently favor a view that describes the N-terminal domain as a large flexible loop, anchored onto the membrane at the amino terminus via the membrane-associated helix and at the C terminus via the first TM. This view is also supported by the recent crystal structures of the b-adrenergic receptor in which the N-terminal domain is so flexible that electron density in this part could not be traced[15, 16]. We have now initiated work on constructs that include parts of the TM bundle to see whether conformational preferences of N-Y4 are influenced by the remainder of the receptor.

5.4 Materials and methods

Expression of the N-Y4 sequence as a soluble fusion to ubiquitin resulted in heterogeneous fragmentation. In order to prevent *in-vivo* processing the N-terminal domain was fused to the highly insoluble protein ketosteroidisomerase that is encoded in the commercial plasmid pET 31b, from which it was liberated by cleavage with the TEV protease in mild detergent.

5.4.1 Plasmid construction, expression and purification of N-Y4

The cDNA of the Y4 receptor was obtained from the University of Missouri-Rolla (UMR) cDNA Resource Center. The following two primers were used to amplify the cDNA corresponding to N-Y4 by PCR. Forward primer: GCG*CTCGAG***GGTTCCGGTTCCGGTTCC**GAAAACCTGTACTTCCAGATGAAC ACCTCTCACCTGCTGGC, in which italic letters denote a XhoI cleavage site, bold letters denote a Gly-Ser linker sequence and underlined letters identify a TEV cleavage sequence; backward primer: CTG*GCTGAGC*TCACATCACGTCCACGGAATCCT with italic letters denoting an EspI cleavage site. The amplified PCR product and the target vector, pET 31b (Novagen), were simultaneously digested with XhoI and EspI, and ligated into the vector with T4 ligase. The construct was confirmed by DNA sequencing (Synergene Biotech, Switzerland). All mutants were constructed by site-directed mutagenesis using the QuikChange Kit (Stratagene, USA).

The fusion protein was expressed in inclusion bodies using the BL21(DE3) *E.coli* strain. Protein expression was performed by growing cells at 37°C using minimal media containing ^{15}N-NH_4Cl as the sole nitrogen source for ^{15}N labeled peptide. 1mM IPTG was added to induce protein expression when the OD_{600} reached 0.8 and cells were harvested after 5-6 hours. The fusion protein was purified from inclusion bodies in 6M guanidinium hydrochloride by Ni-NTA affinity chromatography. After removal of GdnHCl by dialysis the precipitated fusion protein was solubilized in 50mM Tris pH 8.0 in the presence of 2% N-lauryl sarcosine upon sonication to a final concentration of 2mg/ml. The resulting solution was dialyzed against a 20-fold excess of 50 mM Tris, pH 8.0 for 4-6 times. The solution was diluted 10 times with 50 mM Tris pH 8.0 and EDTA and DTT were added to a final concentration of 0.5 mM and 1 mM, respectively. TEV protease was added to a final concentration of 100 mM and the cleavage mixture was kept at 4°C over night. The target peptide was purified by C18-RP-HPLC (Vydac, USA) and the correctness of the peptide was verified by MALDI-TOF MS: ^{15}N labeled N-Y4: 4614 Da (theoretical mass (for 100% labeling): 4611.1 Da).

5.4.2 Synthesis and purification of the neuropeptides and of unlabelled N-terminal fragments

^{15}N-labeled peptides from the NPY family were expressed as soluble fusions to Ubiquitin. Ubiquitin was liberated from the neuropeptide using the yeast ubiquitin hydrolase, and C-terminal amidation was performed using the a-amidating peptidyl glycine amidase (PAM). We have used the protocols for expression, ubiquitin cleavage and C-terminal amidation many times before and described in them much detail elsewhere, e.g. in Bader et al[11].

Wild-type and mutant N-Y4 peptides and peptides from the NPY family containing ^{15}N nuclei at natural abundance were prepared by solid-phase peptide synthesis using a robot system (ABI433A, Applied Biosystems). 2-chlorotrityl chloride resin preloaded with Fmoc-Met-OH was used to assemble the linear peptide using standard Fmoc chemistry (20% piperidine in DMF for Fmoc deprotection, 4 equiv. HOBt/HBTU for activation, diisopropylethylamine as base, and *N*-methylpyrrolidone as solvent). The peptides were cleaved from the resin and deprotected with TFA/water/1,2-ethanedithiol/triisopropylsilane 95/2.5/2.5/2.5. The product was lyophilized and purified by C18 RP-HPLC and correctness was confirmed by ESI-MS: wild-type N-Y4: 4556.8 Da (theoretical mass: 4556.1 Da); K13A N-Y4: 4501 Da (theoretical mass: 4499 Da); R20A N-Y4: 4473 Da (theoretical mass: 4471 Da); K22A N-Y4: 4501 Da (theoretical mass: 4499 Da).

In order to synthesize the N-terminally biotinylated forms the peptides were mixed with biotin-$(PEO)_4$-NHS-propionate (Molecular Biosciences, USA) in a 1:2 ratio in 100mM phosphate buffer, pH 7.0 and incubated for 2 hours at RT and afterwards purified by C18 RP-HPLC and confirmed by ESI-MS. To confirm that in case of E4K-bPP, the biotin was coupled to the N-terminus instead of the side chain of lysine, the biotinylated peptide was first digested with pepsin, and subsequently the fragment containing residue 1-16 was analyzed by MALDI-TOF MS-MS. The result from this analysis demonstrated that the biotin was exclusively coupled to the N terminus.

5.4.3 Dodecylphosphoethanolamine coupling to the carboxyl terminus of N-Y4

The peptide from solid-phase peptide synthesis was cleaved off the resin with TFA (0.8 vol %) in DCM with all the protecting groups remaining intact. Following removal of solvents the protected peptide was precipitated in cold water, lyophilized and redissolved in DMF. The solution was stirred at RT for 5 hours with 3 equivalents of dodecylphosphoethanolamine (3 equiv.) in presence of of HATU (1 equiv.), HOAt (1 equiv.) and of DIEA(1.5 equiv). After extraction with a ethyl-acetate:water mixture (1:1 v/v) the lipopeptide was deprotected under the same conditions as described above. Finally, the lipopeptide was purified by C4 RP-HPLC (Vydac, USA), lyophilized and purity higher than 95% was confirmed by MALDI-TOF-MS: 4848 Da (theoretical mass: 4847.1 Da) and LC-MS.

5.4.4 NMR experiments

All samples of N-Y4 for structural studies were measured at 1mM concentration, 40mM d-MES at pH 5.6. For measurements mimicking membrane environments 300mM d_{38}-DPC or 300mM d_{25}-SDS were added. All experiments were performed at 700 MHz, 310K using a triple-resonance cryoprobe. Resonance assignments were initially performed in the absence of DPC or SDS using [^{15}N,^{1}H]-HSQC, 3D [^{15}N,^{1}H]-HSQC-TOCSY (80ms mixing time) and 300ms 3D [^{15}N,^{1}H]-HSQC-NOESY experiments. Details of the spectroscopy were similar to those described by us earlier[28]. Spectra were analyzed using the programs CARA[29] and XEASY[30]. After nearly complete resonance assignments in water were obtained, a 200ms 3D [^{15}N,^{1}H]-HSQC-NOESY was recorded in the presence of DPC, and the assignments in water adjusted to the DPC spectra. Upper distance restraints in DPC or SDS were then derived from 50ms 2D NOESY spectra. Internal backbone dynamics were studied by measuring a ^{1}H-detected version of a ^{15}N{^{1}H}-NOE experiment. Structures were computed based on upper-distance restraints derived from the NOESY spectra using the program CYANA[31, 32] following the standard simulated annealing protocol. ^{15}N{^{1}H}-NOEs were computed from the ratio of integrals from signals in the presence to those in the absence of amide proton irradiation[33]. Chemical shifts of the ^{15}N,^{1}H-correlation map in the absence and

full assignments in the presence of DPC and SDS can be found in the Supp. Mat. Proton chemical shifts were referenced to the water line, taken as 4.63 ppm at 310K, from which the nitrogen scale was derived indirectly through multiplication with the factor $g(^{15}N)/g(^{1}H)$. The coordinates, chemical shift values and heteronucelar NOEs of NY-4 in the presence of SDS and DPC have been deposited in the BMRB database under the accession number 15708.

5.4.5 Membrane-association topology using spin labels

In the spin label studies [^{15}N,^{1}H]-HSQC spectra of 0.5mM solutions of ^{15}N-N-Y4 containing 300mM DPC were measured in absence and presence of 7mM and 8.8mM 5-doxyl and 16-doxyl stearic acid, respectively. Signal attenuation was computed from the ratio of integrals from peaks in the corresponding spectra. The signal attenuation in the presence of the spin label is related to proximity of protons to the label. In another set of experiments 0.1mM ^{15}N-labeled N-Y4 was mixed with various concentrations of bPP in order to test whether N-Y4 is released from the micelle upon interaction with PP.

5.4.6 Surface plasmon resonance (SPR) studies

HBS buffer (10mM HEPES, pH 7.4, 150mM NaCl, 3.4mM EDTA, 0.005% P20) was used as the running buffer to achieve physiological pH. N-terminally biotinylated neuropeptides were immobilized onto the sensor chip SA (Biacore, Sweden), which contains a streptavidin-coated surface, resulting in about 200 response units (RU) on a Biacore 1000 instrument (Biacore, Sweden). Different concentrations of N-Y4 spanning a range of 5 to 100μM were applied to the surface for 30 seconds at a flow-rate of 20ml/min at 25°C. After each injection of analytes, the flow-cell was flushed with regeneration buffer (1M NaCl, 50mM NaOH) for 30 seconds. Since unspecific binding at concentrations higher than 100μM occurred, K_D larger than 100μM could not be determined precisely. Nevertheless, trends in reduction of binding could still be computed from a limited set of data points, in which values at high concentrations were excluded from the analysis. All sensograms were analyzed with the BIA evaluation software using a two-state binding model.

5.5 Acknowledgements

We like to deeply acknowledge valuable discussion with F. Naider and R. Bader. We would like to thank for financial support from the Swiss National Science Foundation (grant No. 3100A0-11173).

5.6 References

1. Holliday, N.D. and H.M. Cox, *Control of signalling efficacy by palmitoylation of the rat Y1 receptor.* Br J Pharmacol, 2003. **139**(3): p. 501-12.
2. Kalra, S.P. and P.S. Kalra, *NPY and cohorts in regulating appetite, obesity and metabolic syndrome: beneficial effects of gene therapy.* Neuropeptides, 2004. **38**(4): p. 201-11.
3. Larhammar, D., et al., *Origins of the many NPY-family receptors in mammals.* Peptides, 2001. **22**(3): p. 295-307.
4. Berglund, M.M., et al., *Studies of the human, rat, and guinea pig Y4 receptors using neuropeptide Y analogues and two distinct radioligands.* Peptides, 2001. **22**(3): p. 351-6.
5. Lundell, I., et al., *Cloning of a human receptor of the NPY receptor family with high affinity for pancreatic polypeptide and peptide YY.* J Biol Chem, 1995. **270**(49): p. 29123-8.
6. Lundell, I., et al., *The cloned rat pancreatic polypeptide receptor exhibits profound differences to the orthologous receptor.* Proc Natl Acad Sci U S A, 1996. **93**(10): p. 5111-5.
7. Kapust, R.B., et al., *Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency.* Protein Eng, 2001. **14**(12): p. 993-1000.
8. van den Berg, S., et al., *Improved solubility of TEV protease by directed evolution.* J Biotechnol, 2006. **121**(3): p. 291-8.
9. Kapust, R.B., et al., *The P1' specificity of tobacco etch virus protease.* Biochem Biophys Res Commun, 2002. **294**(5): p. 949-55.
10. Papavoine, C.H., et al., *NMR studies of the major coat protein of bacteriophage M13. Structural information of gVIIIp in dodecylphosphocholine micelles.* Eur. J. Biochem., 1995. **232**(2): p. 490-500.
11. Bader, R., et al., *Structure and Dynamics of Micelle-bound Neuropeptide Y: Comparison with unligated NPY and Implications for Receptor Selection.* J. Mol. Biol., 2001. **305**: p. 307-392.
12. Respondek, M., et al., *Mapping the orientation of helices in micelle-bound peptides by paramagnetic relaxation waves.* J Am Chem Soc, 2007. **129**(16): p. 5228-34.
13. Conlon, J.M., *The origin and evolution of peptide YY (PYY) and pancreatic polypeptide (PP).* Peptides, 2002. **23**(2): p. 269-78.

14. Wimley, W.C., T.P. Creamer, and S.H. White, *Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides.* Biochemistry, 1996. **35**(16): p. 5109-24.
15. Rasmussen, S.G.F., et al., *Crystal structure of the human beta(2) adrenergic G-protein-coupled receptor.* Nature, 2007. **450**: p. 383-U4.
16. Cherezov, V., et al., *High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor.* Science, 2007. **318**(5854): p. 1258-65.
17. Kaiser, E.T. and F.J. Kezdy, *Secondary structures of proteins and peptides in amphiphilic environments.* Proc. Natl. Acad. Sci. U S A, 1983. **80**(4): p. 1137-43.
18. Kaiser, E.T. and F.J. Kezdy, *Amphiphilic secondary structure: design of peptide hormones.* Science, 1984. **223**(4633): p. 249-55.
19. Sargent, D.F. and R. Schwyzer, *Membrane lipid phase as catalyst for peptide-receptor interactions.* Proc. Natl Acad. Sci. U S A, 1986. **83**(16): p. 5774-8.
20. Schwyzer, R., *100 Years Lock-and-Key Concept: Are Petide Keys Shaped and Guided to Their Receptors by the Target Cell Membrane?* Biopolymers, 1994. **37**: p. 5-16.
21. Lerch, M., M. Mayrhofer, and O. Zerbe, *Structural similarities of micelle-bound peptide YY (PYY) and neuropeptide Y (NPY) are related to their affinity profiles at the Y receptors.* J. Mol. Biol., 2004. **339**(5): p. 1153-68.
22. Lerch, M., et al., *Bovine pancreatic polypeptide (bPP) undergoes significant changes in conformation and dynamics upon binding to DPC micelles.* J. Mol. Biol., 2002. **322**(5): p. 1117-33.
23. Bader, R. and O. Zerbe, *Are hormones from the neuropeptide Y family recognized by their receptors from the membrane-bound state?* ChemBioChem, 2005. **6**(9): p. 1520-34.
24. Larhammar, D., *Evolution of neuropeptide Y, peptide YY and pancreatic polypeptide.* Regul. Pept., 1996. **62**(1): p. 1-11.
25. Walker, P., et al., *Acidic residues in extracellular loops of the human Y1 neuropeptide Y receptor are essential for ligand binding.* J. Biol. Chem., 1994. **269**(4): p. 2863-9.
26. Merten, N., et al., *Receptor subtype-specific docking of Asp6.59 with C-terminal arginine residues in Y receptor ligands.* J Biol Chem, 2007. **282**(10): p. 7543-51.
27. Lerch, M., et al., *Strongly Altered Receptor Binding Properties in PP and NPY Chimera are Accompanied by Changes in Structure and Membrane Binding.* Biochemistry, 2005. **44**: p. 9255 - 9264.
28. Bader, R., *Structural Aspects of Neuropeptide Y: Implications of the membrane-bound State for Receptor Recognition and Subtype Selectivity Studied by NMR*, in *Institute of Pharmaceutical Sciences, Department of Applied BioSciences*. 2001, ETH Zurich: Zurich. p. 1-234.
29. Keller, R., *The Computer Aided Resonance Assignment*. 2004, Goldau: CANTINA Verlag.
30. Bartels, C., et al., *The program XEASY for computer-supported spectral analysis of biological macromolecules.* J. Biomol. NMR, 1995. **6**: p. 1-10.
31. Guntert, P., *Automated NMR structure calculation with CYANA.* Methods Mol Biol, 2004. **278**: p. 353-78.

32. Guntert, P., C. Mumenthaler, and K. Wuthrich, *Torsion angle dynamics for NMR structure calculation with the new program DYANA*. J Mol Biol, 1997. **273**(1): p. 283-98.
33. Noggle, J.H. and R.E. Schirmer, *The Nuclear Overhauser Effect - Chemical Applications*. 1971, New York: Academic Press.

5.7 Supplementary materials

Fig. S1: "Snake"-plot type presentation of the human Y4 receptor. The plot was downloaded from the GPCR.org website. Note that to enhance clarity not all residues from the N-terminal domain and the long loops are shown

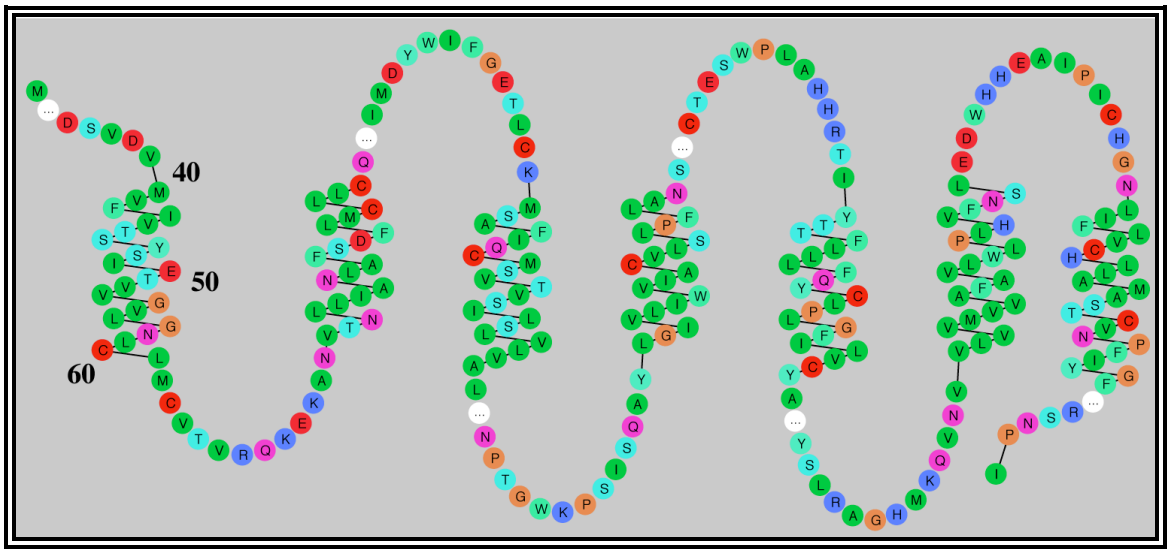


Fig. S2: Summary of the meaningful distance restraints as derived from the unambiguously assigned inter-residue NOEs between backbone H^N,H^α and H^β of N-Y4 bound to DPC micelles.

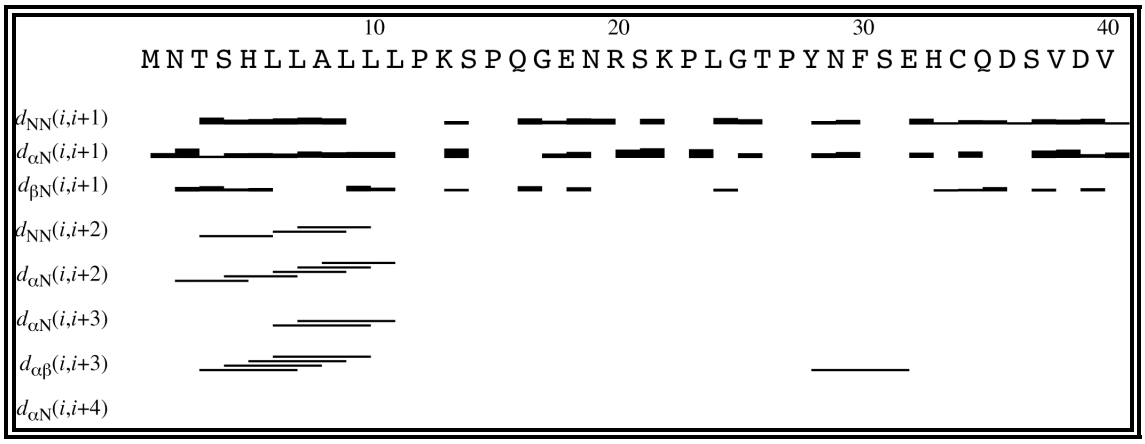


Fig. S3: Expansion of the 100ms NOESY displaying the region involving NOEs between sequential amide protons (DPC micelles)

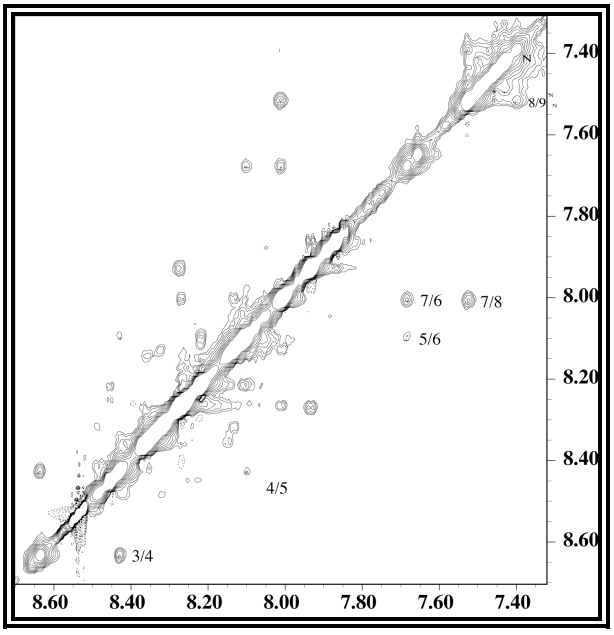


Table S4: Chemical shifts for N-Y4 bound to DPC micelles [a] referenced to the signal of residual HDO at 4.63 ppm

Residue	H^N	H^α	H^β	others
Met 1	-	4.07	-, -	γCH_2 -, -; εCH_3 -
Asn 2	8.99	4.92	2.92, 2.80	δNH_2 7.66, 6.90
Thr 3	8.63	3.89	4.20	γCH_3 1.20; γOH -
Ser 4	8.43	4.14	3.91, 3.91	γOH -
His 5	8.11	4.50	3.25, 3.25	$\delta^1 NH$ -; $\delta^2 H$ 6.89; $\varepsilon^1 H$ 7.66; $\varepsilon^2 NH$ -
Leu 6	7.69	3.98	1.74, 1.74	γH 1.67; δCH_3 0.90, 0.84
Leu 7	8.01	3.91	1.75, 1.75	γH 1.58; δCH_3 0.85, 0.85
Ala 8	7.52	4.03	1.44	
Leu 9	7.40	4.09	1.89, 1.89	γH 1.47; δCH_3 0.83, 0.83
Leu 10	7.43	4.20	1.72, 1.72	γH 1.55; δCH_3 0.83, 0.79
Leu 11	7.50	4.51	1.65, 1.65	γH 1.50; δCH_3 0.87, 0.87
Pro 12		-	2.29, 1.86	γCH_2 1.98, 1.98; δCH_2 3.54, 3.69
Lys 13	8.35	4.30	1.70, 1.70	γCH_2 1.42, 1.42; δCH_2 1.78, 1.78; εCH_2 2.96, 2.96; ζNH_3^+ -
Ser 14	8.30	4.68	3.81, 3.81	γOH -
Pro 15		4.40	2.24, 1.88	γCH_2 1.97, 1.97; δCH_2 3.69, 3.79
Gln 16	8.37	4.27	2.08, 1.94	γCH_2 2.33, 2.33; εNH_2 7.45, 6.78
Gly 17	8.27	3.90, 3.90		
Glu 18	8.22	4.26	2.01, 1.88	γCH_2 2.24, 2.24; εH -
Asn 19	8.45	4.64	2.79, 2.71	δNH_2 7.53, 6.86
Arg 20	8.26	4.31	1.85, 1.71	γCH_2 1.58, 1.58; δCH_2 3.14, 3.14; εNH -; ζNH_2 -,
Ser 21	8.22	4.37	3.80, 3.80	γOH -
Lys 22	8.10	4.57	1.65, 1.65	γCH_2 1.76, 1.76; δCH_2 1.41, 1.41; εCH_2 2.95, 2.95; ζNH_3^+ -
Pro 23		4.40	2.27, 2.27	γCH_2 1.98, 1.98; δCH_2 3.55, 3.55
Leu 24	8.34	4.26	1.63, 1.63	γH 1.55; δCH_3 0.86, 0.86
Gly 25	8.27	3.92, 3.92		
Thr 26	7.94	4.53	4.08	γCH_3 1.15; γOH -
Pro 27		-	2.47, 2.14	γCH_2 1.96, 1.82; δCH_2 3.75, 3.57
Tyr 28	7.91	4.34	2.79, 2.79	δH 6.92, 6.92; εH 6.71, 6.71; ζOH -
Asn 29	8.08	4.61	2.71, 2.59	δNH_2 7.48, 6.79
Phe 30	8.21	4.44	3.13, 3.05	δH 7.24, 7.24; εH 7.14, 7.14; ζH -
Ser 31	8.07	4.39	3.81, 3.81	γOH -
Glu 32	8.08	4.16	1.94, 1.85	γCH_2 2.21, 2.21; εH -
His 33	8.24	4.59	3.28, 3.09	$\delta^1 NH$ -; $\delta^2 H$ -; $\varepsilon^1 H$ -; $\varepsilon^2 NH$ -
Cys 34	8.20	4.39	2.86, 2.86	γSH -
Gln 35	8.43	4.25	2.08, 1.96	γCH_2 2.31, 2.31; εNH_2 7.43, 6.76
Asp 36	8.25	4.64	2.76, 2.64	δH -
Ser 37	8.13	4.42	3.80, 3.80	γOH -
Val 38	8.01	4.11	2.06	γCH_3 0.86, 0.86
Asp 39	8.27	4.26	2.74, 2.60	δH -
Val 40	7.94	4.09	2.09	γCH_3 0.87, 0.87
Met 41	7.87	4.26	1.94, 1.94	γCH_2 -, -; εCH_3 2.04

[a] 1mM in 300mM DPC / 90% H_2O/10% ^2H_2O at 310 K , 20mM MES and pH 5.6

Table S5: $^{1}H^{N}$ and ^{15}N chemical shifts of N-Y4 in water in the presence and absence of DPC micelles.

Res.	δ(^{1}H) [ppm]	δ(^{15}N) [ppm]	δ(^{1}H) [ppm] DPC	δ(^{15}N) [ppm] DPC
M1	-	-	-	-
N2	-	-	-	-
T3	8.171	119.838	8.637	116.741
S4	8.290	114.787	8.413	117.173
H5	8.260	127.348	-	-
L6	-	-	7.652	118.535
L7	8.006	121.818	8.000	116.495
A8	7.953	123.247	7.521	118.535
L9	7.817	119.790	7.395	115.914
L10	7.915	121.832	7.422	115.509
L11	7.897	123.675	7.494	117.061
P12	-	-	-	-
K13	8.257	120.984	8.350	120.579
S14	8.221	117.544	8.319	118.053
P15	-	-	-	-
Q16	8.338	119.671	8.355	119.671
G17	8.243	109.457	8.255	109.367
E18	8.225	120.207	8.232	120.105
N19	8.454	119.731	8.449	119.615
R20	8.252	121.321	8.247	121.173
S21	8.211	116.370	8.214	116.299
K22	8.092	123.461	8.095	123.333
P23	-	-	-	-
L24	8.423	122.478	8.319	121.952
G25	8.315	109.153	8.286	108.684
T26	7.874	115.514	7.910	115.127
P27	-	-	-	-
Y28	-	-	7.886	118.731
N29	8.083	120.162	8.045	119.607
F30	7.989	120.843	8.083	120.695
S31	8.069	116.045	8.084	115.520
E32	8.133	121.685	8.162	121.566
H33	8.288	117.461	8.332	117.816
C34	-	-	-	-
Q35	8.430	121.020	8.460	121.068
D36	8.224	120.984	8.289	121.269
S37	8.096	115.276	8.123	115.324
V38	8.002	120.362	8.008	120.111
D39	8.251	122.983	8.247	122.843
V40	7.918	119.379	7.925	119.118
M41	7.850	128.423	7.831	127.626

Fig. S6: Summary of the meaningful distance restraints as derived from the unambiguously assigned inter-residue NOEs between backbone H^N,H^α and H^β of N-Y4 bound to SDS micelles.

Fig. S7: Comparison of [^{15}N,^{1}H]-HSQC data of N-Y4 in the presence of DPC (left) or SDS (right) micelles

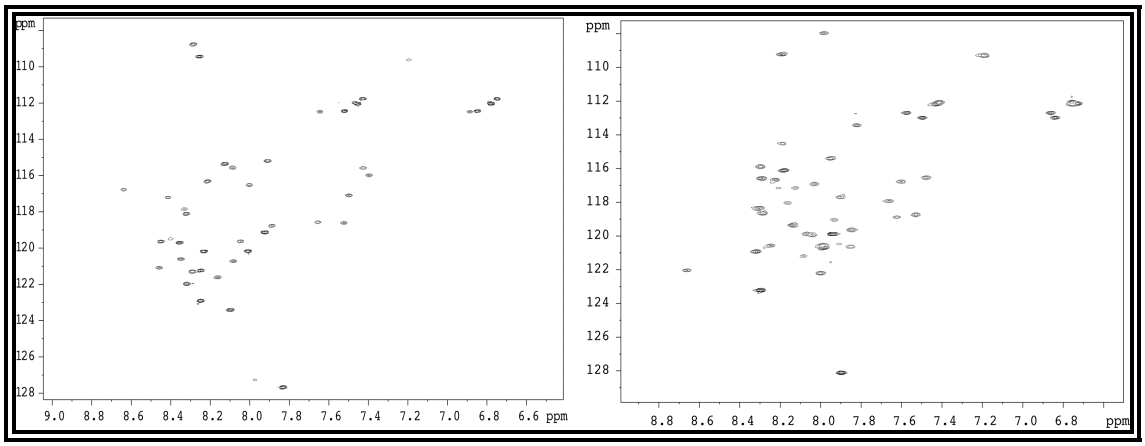


Fig. S8: Comparison of $^{15}N\{^{1}H\}$-NOE data of N-Y4 in the presence of DPC (red diamonds) or SDS (black circles) micelles.

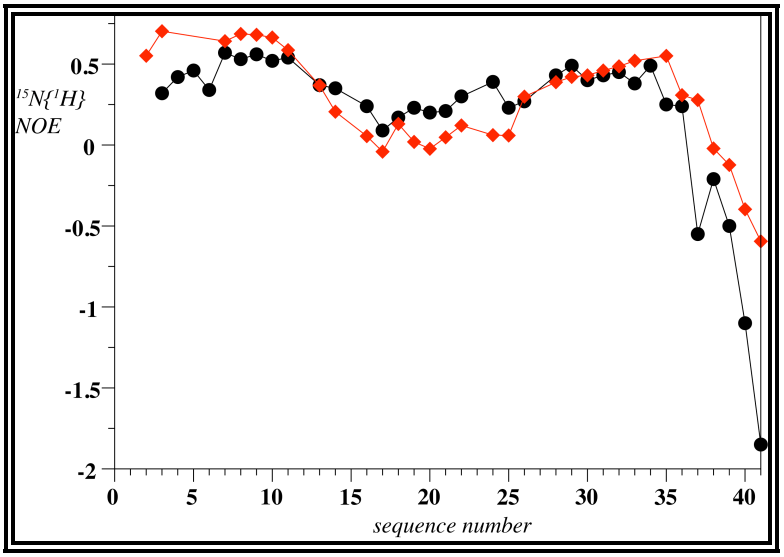


Fig. S9: Left: Residual signal intensity of cross peaks in the [^{15}N,^{1}H]-HSQC in the presence of 5-doxylstearate relative to those in the absence of the spinlabel. Right: Free energies of transfer for whole amino acids from bulk aqueous solution into the water-membrane interface. Values were taken from Wimley et al[36]

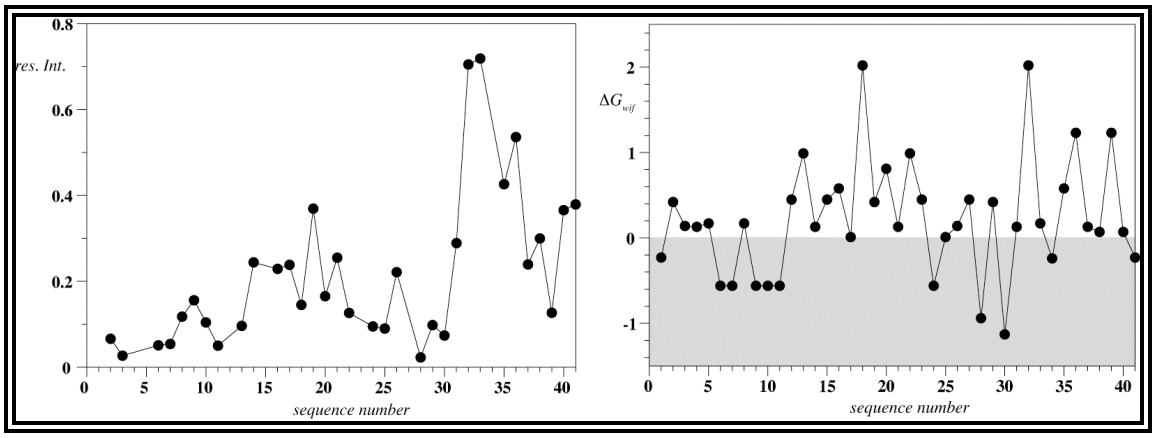


Fig. S10: Left: Overlay of NOESY spectra of N-Y4 with the construct containing the C-terminal lipid attachment (green) (all spectra recorded in the presence of DPC micelles). Right: Expansion showing the assignment of the helical segment of N-Y4.

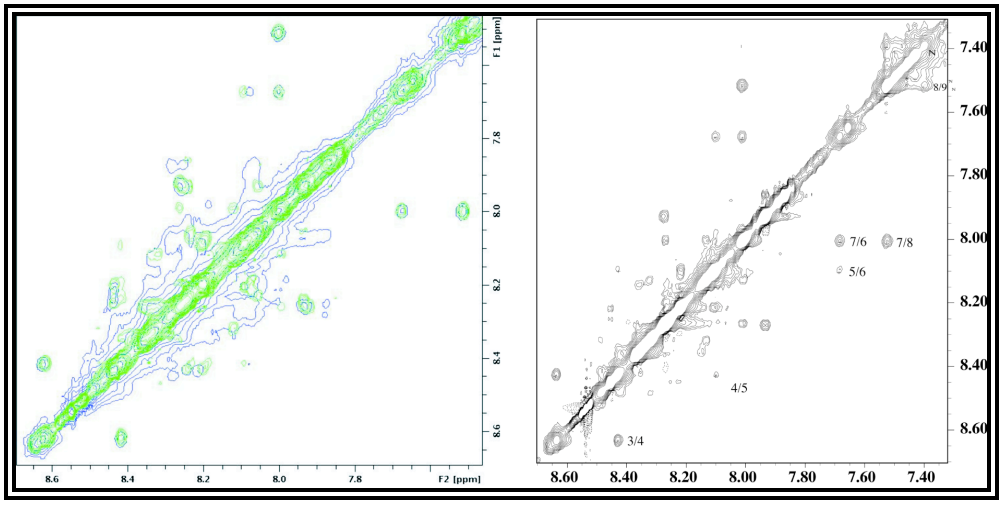


Table S11: Chemical shifts for N-Y4 bound to SDS micelles[a] referenced to the signal of residual HDO at 4.63 ppm

Residue	H^N	H^α	H^b	Others
Met 1	-	4.19	2.19	γCH_2 2.55; εCH_3 2.19
Asn 2	8.66	4.98	2.98 , 2.79	δNH_2 7.57, 6.86
Thr 3	8.28	4.26	4.04	γCH_3 1.22; γOH -
Ser 4	8.28	4.23	3.92	γOH -
His 5	8.06	4.57	3.3	δ^1NH -; δ^2H 7.34; ε^1H 8.64; ε^2NH -
Leu 6	7.84	3.99	1.69, 1.78	γH 1.57; CH_3 0.87, 0.94
Leu 7	8.03	3.95	1.57, 1.78	γH 1.58; δCH_3 0.87, 0.93
Ala 8	7.52	4.05	1.47	
Leu 9	7.47	4.14	1.70, 1.87	γH 1.54; δCH_3 0.86, 0.93
Leu 10	7.59	4.22	1.75, 1.80	γH 1.55; δCH_3 0.82, 0.86
Leu 11	7.65	4.37	1.77	γH 1.56; δCH_3 0.90, 0.94
Pro 12		4.45	2.35, 2.05	γCH_2 2.0, δCH_2 3.54, 3.77
Lys 13	7.90	4.44	1.81	γCH_2 1.46; δCH_2 1.63 ; εCH_2 - ; ζNH_3^+ 6.96
Ser 14	7.95	3.88	3.80	γOH -
Pro 15		4.46	2.31, 1.99	γCH_2 2.06; δCH_2 3.80, 3.75
Gln 16	8.30	4.31	2.15, 1.97	γCH_2 2.30, 2.37; εNH_2 7.42, 6.72
Gly 17	8.18	3.93		
Glu 18	8.14	4.30	2.08, 1.94	γCH_2 2.33; εH -
Asn 19	8.30	4.71	2.84, 2.71	δNH_2 7.49, 6.83
Arg 20	8.0	4.35	1.86, 1.78	γCH_2 1.64; δCH_2 3.20; εNH 7.2; ζNH_2 -
Ser 21	8.22	4.43	3.83	γOH -
Lys 22	7.99	4.35	1.87	γCH_2 1.45; δCH_2 1.75 ; εCH_2-; ζNH_3^+ -
Pro 23		4.44	2.30, 1.98	γCH_2 1.91; δCH_2 3.77, 3.68
Leu 24	8.03	4.34	1.72, 1.68	γH 1.58; δCH_3 0.94, 0.89
Gly 25	7.98	4.00, 3.91		
Thr 26	7.82	4.61	4.17	γCH_3 1.21; γOH -
Pro 27		4.39	2.17	γCH_2 2.17; δCH_2 3.58
Tyr 28	7.61	4.34	2.71, 2.67	δH 6.88; εH 6.71; ζOH -
Asn 29	7.93	4.68	2.80, 2.63	δNH_2 7.42, 6.76
Phe 30	8.08	4.43	3.20, 3.07	δH 7.28; εH 7.21; ζH 7.07
Ser 31	8.18	4.31	3.93, 3.92	γOH -
Glu 32	7.85	4.23	2.03, 1.91	γCH_2 2.23; εH -
His 33	8.12	4.62	3.34, 3.20	δ^1NH -; δ^2H7.31; ε^1H -; ε^2NH -
Cys 34	8.15	4.66	3.22, 2.98	γSH -
Gln 35	8.24	4.33	2.11, 1.98	γCH_2 2.34; εNH_2 7.41, 6.75
Asp 36	8.31	4.65	2.75, 2.69	δH -
Ser 37	8.18	4.45	3.85	γOH -
Val 38	8.00	4.14	2.09	γCH_3 0.89, 0.85
Asp 39	8.30	4.62	2.60	δH -
Val 40	7.94	4.12	2.09	γCH_3 0.90
Met 41	7.89	4.34	1.92	γCH_2 2.34; εCH_3 -

[a]1mM in 300mM SDS / 90% H_2O/10% ^2H_2O at 310 K, 20mM MES and pH 5.6

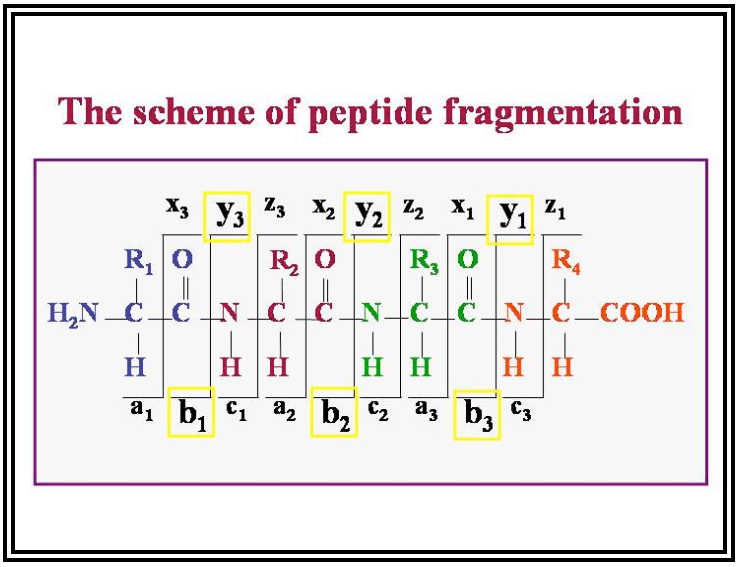
Selective N-terminal biotinylation: Proof by MS-MS

A P L K P E Y P G D N A T P E QMA Q Y A A E L R R Y I N M L T R P R Y-NH_2

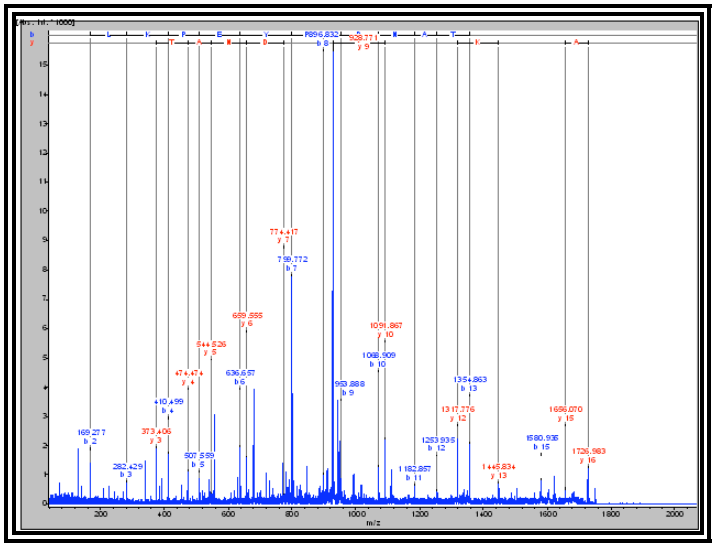
MW=4224.8 Da

Biotin-(PEO)4-NHS-propionate (NHS-(PEO)4-Biotin): MW=588.67 Da, mass added to the target: 474.6.Da

Pepsin was used to digest the biotinylated peptide yielding a fragment corresponding to biotinylated 1-16 N-Y4. TOF-TOF MS was utilized to analyze this fragment:

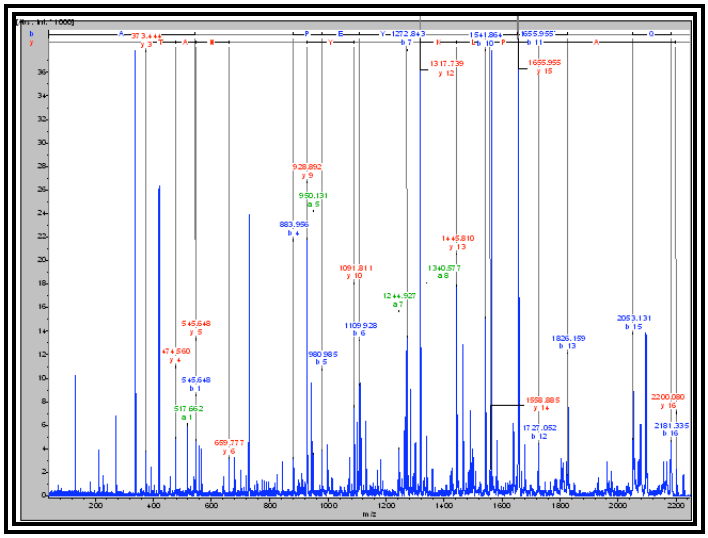


The non-biotinylated sample served as a control, and the TOF-TOF result is as follows:



	y3	y4	y5	y6	y7	y8	y9	y10	y11	y12	y13	y14	y15	y16
Expected	372.3	473.4	544.5	658.6	773.7	830.8	927.9	1091.1	1220	1317.3	1445.5	1558.6	1655.7	1726.8
Observed	373.4	474.5	544.5	659.5	774.4		928.8	1091.9		1317.8	1445.8		1656.1	1726.9

Almost all expected ions from the digest are observed. Thereafter, the biotinylated sample was fragmented under the same conditions, and the TOF-TOF MS spectrum is depicted below:



If we assume that biotin was coupled to the N terminus, the following expected y ions as listed

	y4	y5	y6	y7	y8	y9	y10	y11	y12	y13	y14	y15	y16
Expected[1]	473.4	544.5	658.6	773.7	830.8	927.9	1091.1	1220	1317.3	1445.5	1558.6	1655.7	2201.1
Expected[2]	473.4	544.5	658.6	773.7	830.8	927.9	1091.1	1220	1317.3	1920.1	2033.2	2130.3	2201.1
Observed	474.6	545.6	659.8			928.9	1091.8		1317.7	1445.8	1558.9	1656.1	2200.1

Expected[1]: expected y ions if biotin was coupled to the N terminus
Expected[2]: expected y ions if biotin was coupled to lysine

The observed and expected y ions have been summarized in the table above. It is evident that the observed y ions, in particular y ions 13-16 correspond to those where biotin was coupled to the N terminus. In addition b1 ions 545.6 can be identified corresponding to biotinylated N terminus residue (MW of biotinilyated Alanine b ion=546.0). None of the fragments occurring only for Lys-biotinylated peptide was observed.

Chapter 6

Properties of the N-terminal domains from Y receptors probed by NMR spectroscopy

6.0 Introduction

G-protein-coupled receptors (GPCRs), the largest family of membrane proteins, represent ideal therapeutic targets for a number of disorders and diseases[1]. Although significant progress has been made in structural studies of GPCRs, a much more detailed picture still is highly desirable. During the recent years only, we have witnessed a relative explosion in the amount of structural information available for the GPCR family with two new structures of opsin in the presence and absence of transducin peptide[2, 3]four new structures of β-adrenergic receptors [4-7], and a recent structure of the human adenosine A_{2A} receptor[8] and structure of squid rhodopsin which stimulates a Gq-type G protein[9]. These structures open the door to address many questions of the molecular mechanisms of GPCR activation and G-protein coupling and provide initial glimpses of the answers.

GPCRs have a structurally homologous core of seven transmembrane α helices, but the size and structure of their extracellular and intracellular elements vary wildly, with greatest diversity observed in the extracellular N termini, and other regions like the third intracellular loops or C termini[10]. The N-terminus sequence is relatively short (10–50 amino acids) for monoamine and peptide receptors, and much larger (350–600 amino acids) for glycoprotein hormone receptors, and the glutamate family receptors. The largest amino terminal domains are observed in the adhesion family receptors. Mutagenesis and biophysical analysis of several GPCRs indicate that small molecule agonists and antagonists bind to a hydrophobic pocket buried in the transmembrane core of the receptor.

In contrast, peptide ligands bind to both the extracellular and transmembrane domains. The N termini from family 3 GPCRs are the largest among all GPCRs, comprising usually more than 500 amino acids[11]and shares a low but significant sequence similarity with periplasmic binding proteins of bacteria. They constitute of two lobes separated by a hinge region, which looks like a Venus' flytrap upon binding of the ligand. Grafting and mutagenesis studies have demonstrated conserved serine and threonine residues in these domains are directly involved in ligand binding[12]. Surprisingly, the expressed N terminus alone can bind the ligand with affinity similar to the one from the full-length receptor.[13] For the family 2 receptors, which are activated by large peptides like glucagon or secretin, VIP or PACAP, a conserved region in the relatively long N-terminal domain plays a role in the binding of the ligand[11].

In the family 1 GPCRs activated by small ligands like catecholamines (subfamily 1a), the ligands bind in a cavity formed by TM-III to TM-VI without the involvement of N-terminus. Other family 1 GPCRs (subfamily 1b) are activated by short peptides which interact with the extracellular loops and the N terminal domain[14]. In case of the subfamily 1c GPCRs, a large N-terminal extracellular domain recognizes and binds the glycoproteins, and allows them to activate the core domain via its interaction with e1 and e3 loops (Ji and Ji, 1995; Fernandez and Puett, 1996). The N-terminal domains from family 1 GPCRs have received little attention, most likely because of their short length, usually less than 70 amino acids. However, recent studies have suggested a pivotal role of N termini from GPCRs of this class in ligand recognition and binding[15-17]. Furthermore, mutagenesis data highlight the prominent role of charged residues for ligand binding[18, 19]. Koller demonstrated that the N terminus of the calcitonin-like receptor is not only essential for binding to the ligands but also presents a determinant for ligand specificity[20]. The 35 amino-terminal residues of CCR2, expressed as a membrane-bound fusion protein, bind to its ligand with an affinity similar to that of the intact, wild-type receptor, indicating that the N terminus is sufficient for ligand binding in that case[21]. Based on the mutagenesis data on the N terminus of CX3C receptor and previous studies, Chen has proposed a two-step binding model, which comprises ligand binding followed by receptor activation. Therein, the residues located in the N-terminal

domain play distinct roles during the different processes[22]. Complementary to the biological work described above GPCR fragments have been also studied using NMR. For example, Pervushin investigated the N-terminal domain of bacteriorhodopsin, a protein that is structurally highly related to GPCRs, in SDS micelles[23], and Ulfers studied the extracellular domain of the neurokinin-1 receptor in DPC micelles[24]. Riek presented a high-quality 3D NMR structure of the extracellular domain of CRF-R2β in complex with the peptide antagonist astressin[25]. The group of Yeagle has determined conformational preferences for peptides corresponding to the cytosolic loops[26], the 6th TM helix[27] and the N-terminus[26] of rhodopsin. The extracellular N-terminal domain of bovine rhodopsin was surprisingly well-structured and revealed the non-anticipated presence of a short anti-parallel beta-sheet, whereas the corresponding segment of the β -adrenergic receptor could not be traced in the electron maps presumably because of its inherent flexibility[28, 29].

Neuropeptide Y receptors, also called as Y receptors, are members of the rhodopsin-like G-protein coupled receptor (GPCR) family 1b. The neurohormones neuropeptide Y (NPY), peptide YY (PYY) and the pancreatic polypeptide (PP) target a heterologous population of at least five different receptor subtypes Y1, Y2, Y4, Y5 and y6[30]. Their physiological role in the regulation of blood pressure, memory retention, food uptake and seizure has been demonstrated. NPY and PYY posses a similar pharmacology displaying nanomolar affinities for all receptor subtypes[31], whereas PP binds with very high affinity and selectivity to the Y4 receptor[32]. In this work we now report on our recent studies on structural properties of N-terminal domains Y1,Y2,Y4,Y5 the human Y receptors. Synthetic routes for recombinant production of the polypeptides in isotopically-labeled form are described and compared to each other. Expression of N-Y1 and N-Y4 required fusion to the insoluble protein ketosteroidisomerase, from which it was liberated by enzymatic cleavage using the TEV protease in the presence of a mild detergent. In contrast, N-Y2 and N-Y5 could be expressed as soluble ubiquitin fusions, and cleavage was easily achieved with the help of yeast ubiquitin hydrolase.

The N-terminal domains from these Y receptors are fully unstructured in aqueous solution, as shown by measurements of the internal backbone dynamics. In contrast, in the presence of phospholipid micelles N-Y4 adopts a short a-helix in a segment mainly comprised of hydrophobic residues. N-Y1 is largely helical although remaining flexibility precludes a detailed structural analysis. N-Y5 is segregated into more structured and rather flexible regions, similarly to N-Y4. However, measurements of internal backbone dynamics revealed secondary structure to be less stable than in N-Y4. N-Y2 does not interact with the micelles and remains unstructured also in that environment.

6.1 Results

6.1.1 Expression of N-terminal domains in isotopically-labeled form

The structure of peptides can mostly be solved by relying solely on homonuclear ^{1}H-^{1}H correlation experiments. Such peptides are therefore usually produced by solid phase peptide synthesis (SPPS)[33, 34]. Isotope labeling, however, is required for the study of backbone dynamics using ^{15}N relaxation, and such labeling also facilitates chemical shift mapping experiments for the study of macromolecular interactions. The high cost of ^{13}C- and/or ^{15}N-enriched amino acids usually prohibits the usage of SPPS and necessitates recombinant production. *E.coli* is still the expression system of choice for most proteins because of the ease of its genetic manipulation and because of the ability of *E.coli* to synthesize amino acids from glucose and inorganic ammonia salts serving as the sole sources of carbon and nitrogen, respectively[35].

Since peptides are rapidly degraded in *E.coli*, they are usually expressed linked to a more stable fusion partner. The chosen fusion partner should allow the expression of the fusion constructs in high amounts and it should allow the specific separation of the fusion partner and the peptide after the purification of the fusion construct[36]. The first aim can usually be achieved by selecting a protein as fusion partner, which itself can be produced in high yields in *E.coli*. Specific cleavage from the fusion partner can be accomplished for systems for which a specifically hydrolase is available (method 1), or by introducing a unique cleavage site between the fusion partner and the peptide sequence of interest (method 2) (see Fig. 1). Such a scission site can be either an amino

acid sequence specifically recognized by a protease (method 2a) or a site prone to chemical cleavage (method 2b). The most convenient strategy in terms of workload is usually method 2b. In this method high yield is achieved by fusing the peptide to a highly water insoluble protein which will lead to the accumulation of the fusion construct in inclusion bodies[35]. Inclusion bodies already contain the target protein at high concentrations, and typically require only very few additional steps of purification. In case of cyanobromide cleavage very high efficiencies have been reported[37], but the target sequence is not allowed to contain Met residues. Other methods such as hydroxylamine cleavage[38], are much less efficient, and often result in further chemical modifications of the target peptide. Methods that use enzymatic cleavage (methods 1 and 2a) require that the fusion protein is soluble under conditions that are compatible with enzymatic activity.

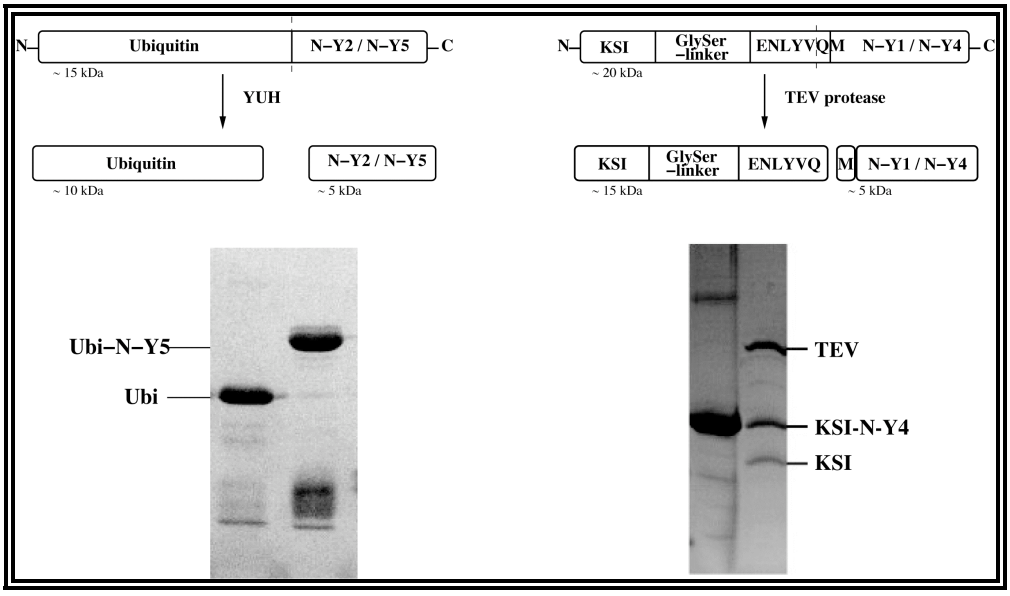


Figure 1: Scheme showing the two strategies used to produce peptides corresponding to the N-terminal domains of the Y receptors in isotope-labeled form.

Since the four Y receptor N-terminal fragments studied herein are all reasonably water-soluble and contain Met residues we initially decided to express them in ^{15}N-labelled form as C-terminal fusions to N-terminally decahistidine-tagged yeast ubiquitin[39]. After purification of the fusion construct by Ni-affinity chromatography the desired peptide was liberated with a hexahistidine-tagged yeast ubiquitin hydrolase

(YUH). This system allowed the recovery of about 6 mg of ^{15}N-labelled N-Y2 and N-Y5 from 1 L of culture. Unfortunately, attempts to express N-Y1 and N-Y4 using this method resulted in unspecific C-terminal degradation (see supplementary Table S10). To circumvent intracellular proteolysis, N-Y1 and N-Y4 were expressed as a fusion to the highly water-insoluble protein ketosteroidisomerase (KSI), which resulted in accumulation of the fusion protein in inclusion bodies. A TEV protease cleavage site was introduced between KSI and the target peptide[40, 41]. The sequence recognized by the TEV protease is ENLYFQ with Q as the P1' residue. To achieve the natural peptide sequence after cleavage, the P1' residue was replaced with the first residue from the target sequence (here it is Met)[40], and an additional GSGSGS linker was inserted to prevent steric hindrance during cleavage.

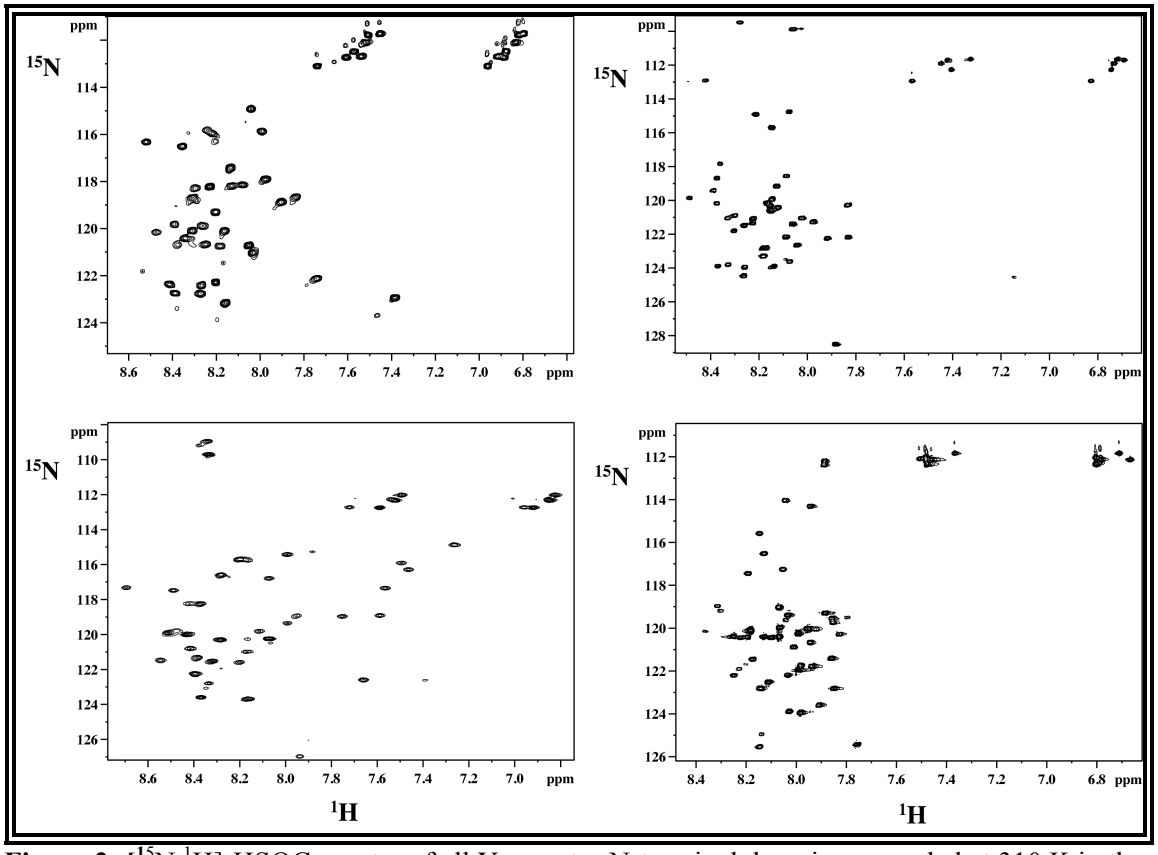


Figure 2: [^{15}N,^{1}H]-HSQC spectra of all Y receptor N-terminal domains, recorded at 310 K in the presence of DPC micelles. Top left: N-Y1, top right: N-Y2, bottom left: N-Y4, bottom right: N-Y5.

A problem of the chosen strategy was that the water-insoluble fusion protein must be solubilized in detergent that is compatible with activity of the TEV protease[42]. After extensive detergent screening, we observed that the ionic detergent sarcosyl solubilizes the fusion protein while preserving TEV protease activity to a satisfactory extent (see Table S10). Cleavage efficiency for this system is around 40% allowing recovery of about 2 mg of ^{15}N-labelled N-Y1 and N-Y4 from 1 L of bacterial culture.

6.1.2 Assignment of chemical shifts

Sequence-specific resonance assignments were done using the strategy developed by Wüthrich and coworkers[43]. In this strategy spin systems are assigned by experiments based on scalar couplings (e.g. COSY-type or TOCSY-type transfer) and NOEs are used to link them in sequential order. Due to extensive resonance overlap of the poorly folded peptides ^{15}N-resolved three-dimensional TOCSY or NOESY data had to be utilized for this task. Representative [^{15}N,^{1}H]-HSQC spectra of all four peptides are depicted in Fig. 2. In case of N-Y5 a ^{13}C,^{15}N-labelled sample, allowing the acquisition of triple resonance spectra, was required. For N-Y1, a set of experiments was first recorded in aqueous buffer. After completed analysis in water the assignments were transferred and adjusted to the spectra recorded in the presence of DPC micelles with the help of NOESY spectra. A set of tables containing complete assignments for all the proton-nitrogen correlation maps, as well as almost all proton chemical shifts of N-Y1 can be found in the supplementary materials S1-S5.

6.1.3 Screening structural properties using ^{15}N relaxation and CD spectroscopy

CD spectroscopy is a convenient tool to estimate the type and content of secondary structure in peptides and proteins. The CD spectra of all N-terminal domains in the presence of DPC micelles are depicted in Fig 3. The spectrum of N-Y2 displays its maximum around 197 nm, the typical absorption band of unstructured peptides. For all other peptides the main band is red-shifted and indicates population of helical substructures. The intensities of the absorptions, however, also clearly show that the helical content is very low in all cases, and the typical bands at 208 and 222 nm are not

visible. For N-Y4, for which we previously observed an a-helix involving residues 4 to 11, the absorption is stronger than for the other peptides.

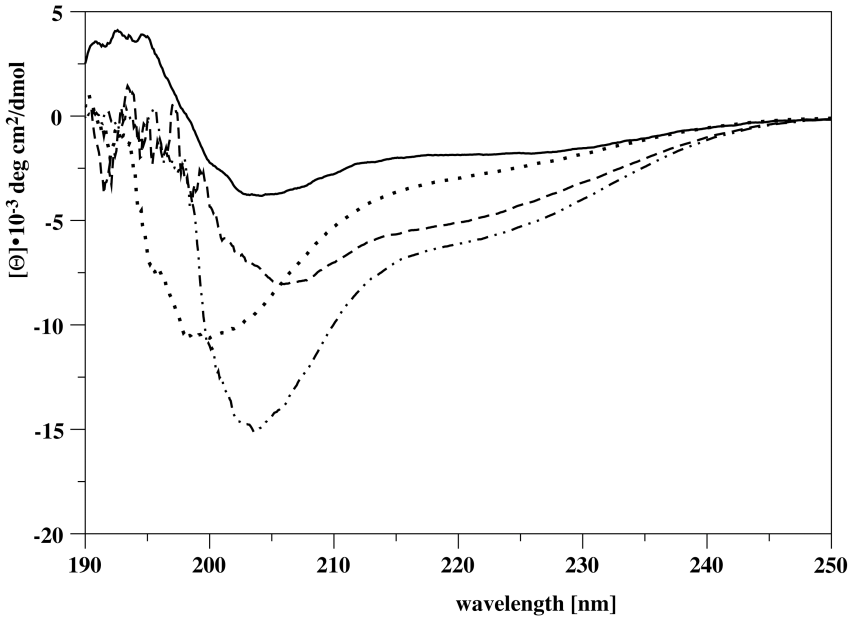


Figure 3: CD spectra of peptides from all N-terminal domains, recorded at 37 °C in 300 mM DPC, 20 mM MES pH 5.6 solution. Data are shown for N-Y1 (solid line), N-Y2 (dotted line), N-Y4 (dash-dotted line) and N-Y5 (dashed line). Data are converted to molar ellipticities.

The dispersion of the NMR signals in the region of the amide protons is traditionally used to estimate to which extent a peptide or protein is folded[44]. In case of the N-terminal domains from the Y receptors signal dispersion of all peptides was small, indicating that they were largely unfolded. To better access whether these peptides still contained folded segments we recorded the ^{1}H-$^{15}N\{^{1}H\}$-NOEs (H-NOEs). These values range from 0.6 and 0.8 for well-folded elements of secondary structures, and progressively decreases for more flexible amide moieties resulting in negative values for fully flexible segments[45]. The H-NOE data for all N-terminal peptides reveal that all peptides are essentially unstructured in aqueous buffer (data not shown).

Since in the naturally occurring GPCR the N termini are attached to a membrane-protein the backbone dynamics were additionally probed in the presence of a commonly used membrane-mimicking detergent, dodecylphosphocholine (DPC)[46] (see Figure 4). Again the peptides are not rigidly structured. In the case of N-Y4 we could previously

show that a rather stable hydrophobic a-helix is formed between residues 4 and 11, present both in zwitterionic (DPC) as well as in anionic (SDS) micelles[47], reflected by H-NOEs exceeding values of 0.6. In contrast, the N termini from all other Y receptors are less well ordered. The N-Y2 is fully flexible most likely due to the complete lack of interactions with phospholipids surfaces. The absence of such contacts is supported by the fact that essentially no chemical shift changes occur between N-Y2 in aqueous buffer and in DPC micelles. In contrast, both N-Y1 and N-Y5 reveal short stretches of the polypeptide chain that become rigidified in the presence of the micelles.

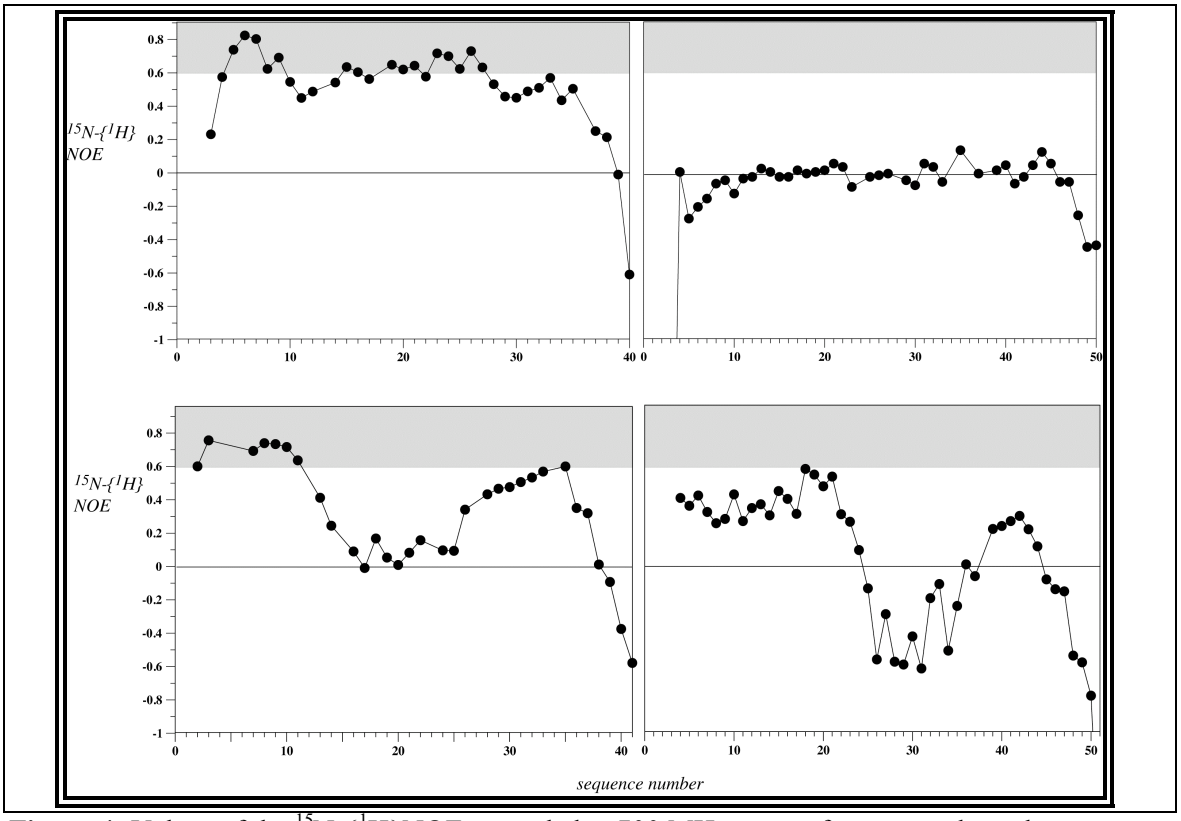


Figure 4: Values of the ^{15}N-{^{1}H}NOE, recorded at 700 MHz proton frequency along the sequence for Y1 (top left), Y2 (top right), Y4 (bottom left) and Y5 (bottom right). The area containing values larger than 0.6, indicating rather well folded segments, has been shaded in gray.

6.1.4 Structures of the N-terminal domains in presence of phospholipid micelles

The NOE data of N-Y4 revealed the presence of a hydrophobic helix in the segment comprising residues 5 to 11. In addition, a nascent helix was observed in the

region including residues 26 to 35. Inspection of the H-NOE data depicted in Figure 4 clearly indicates that N-Y2 is devoid on any structured segments. Moreover, the H-NOE of N-Y5 is generally below 0.6 and mostly values are even smaller than 0.4. In our experience secondary structure cannot reliably be determined in these cases. We speculate that the molecule, similarly to N-Y4, is segregated into a N-terminal helical region, and a much more destabilized shorter C-terminal helical region separated by a longer non-ordered segment, but the peptide is not ordered sufficiently well to allow for structural characterization by NMR in detail.

In case of N-Y1, however, elevated values of the H-NOE are observed indicating that this polypeptide may be amenable to more detailed structural studies. Accordingly, we have assigned all proton and nitrogen resonances of N-Y1. Little chemical shift dispersion of amide proton resonances complicated the assignment process, and use of 3D ^{15}N-resolved NOESY or TOCSY spectra had to be made. During assignment a larger number of contacts involving sequential amide protons were observed, indicating that the f,y space of helical backbone conformations was significantly populated. Such stretches were for example observed involving residues 4 to 9 and residues 24 to 32. An expansion of the spectral region of the [^{1}H,^{1}H]-NOESY that displays the sequential amide proton NOEs in the segment from 24 to 32 is shown in Figure S6 in the supplementary materials. However, except for two αN (i,i+3) NOEs observed in the segment 4-9 no medium-range contacts were found.

The relative strength of intra-residual and sequential αH,NH contacts changes between extended and helical conformations[48], with the intra-residual distance in helices stronger than the sequential one, whereas in extended or unfolded segments the sequential distance is much shorter. A comparison of peak intensities revealed that the sequential NOEs were generally stronger, and in the light of sequential contacts of amide protons, indicate conformational averaging between helical and extended conformations to some extent. Considering this observation it was not really surprising that persistent violations remained in the structure calculations, and helical conformations were only seen involving residues 4 to 9, a region, in which the H-NOE is larger than 0.6. The

$^{3}J(HN,H\alpha)$ couplings were larger than 6.5 Hz throughout the sequence (data not shown), reflecting the remaining conformational instability of N-Y1. To our surprise we have not been able to detect any medium range contacts in the segment 15 to 28, which according to the dynamics data should also be better ordered. We suspect this region to be transiently helical considering the occurrence of sequential amide proton contacts throughout this segment. To summarize, the spectroscopic data indicate that N-Y4 and N-Y5 are similar in that both contain two helical regions separated by a flexible central segment, with only the N-terminal helix in N-Y4 being well ordered. N-Y1 is largely helical between residues 4 and 28, but the remaining conformational flexibility precludes its detailed structural analysis. N-Y2 is fully flexible and devoid of any detectable residual structure.

6.1.5 Interaction studies with neuropeptides from the NPY family

We have recently proposed that the peptides of the NPY family may transiently bind to the N-terminal domains in order to become transferred from the membrane-bound state into the genuine binding pocket of the receptor[47, 49]. While in that work surface plasmon resonance was used to establish the strength of the bPP-NY4 interaction, preliminary experiments using bPP or pPYY and the N-terminal domains from the other receptors have indicated that the interaction between the peptides and the other N-terminal domains are too weak to be detected by SPR. In this work we have now utilized chemical shift mapping experiments both in the presence and absence of DPC micelles in order to derive preliminary data on binding of the peptides from the NPY family to N-Y1, N-Y2 and N-Y5. The changes of chemical shifts of the neuropeptides upon adding 2 equivalents of the N-terminal domains are summarized in Figure 5. First of all we noticed that the magnitude of the changes is larger in the presence of DPC micelles as compared to aqueous buffer. This was surprising considering that in case of N-Y4 the interaction in the absence of micelles was stronger than in the presence as judged by SPR[47]. It should be mentioned, however, that changes in chemical shift mapping depend on both the population difference between bound and free species, but also on the type of structural changes. Contacts with aromatic residues for example will for instance result in larger changes than contacts with polar side chains. Detachment of peptides from the micelle

surface, possibly required for binding to the receptor N termini, will transfer part of the residues into a completely different environment, and we believe this effect to account for the larger changes observed in DPC micelles.

In the presence of lipids, chemical shift changes for bPP are much larger than for the other peptides. When adding N-Y1 or N-Y5 to bPP the same residues of bPP are affected in a similar manner. In all cases the largest changes occur in the N-terminal half of bPP or in the segment from residues 26 to 35. In case of pNPY changes are only appreciable for residues 14 and 15 upon addition of N-Y5. Changes in pPYY are very similar for all the N-terminal domains, and largest for residues 21, 26 and 27.

In the absence of lipids, larger changes are detected for pNPY throughout the sequence upon adding N-Y1 or N-Y2, but are fairly small for bPP and pPYY, except for residue 4 of bPP upon addition of N-Y1. Pronounced differences are observed for pNPY residues 6, 14, 24, 26 and 36, and for residues 4 and 10 when adding N-Y1 or N-Y2, respectively. We like to emphasize here that the spectra of bPP and pPYY in water are concentration independent whereas the oligomerization state of pNPY is strongly concentration dependent[50-52]. In addition, the spectrum of pNPY is highly dependent on pH and other environmental variables. Although we tried to control these as tightly as possible we cannot fully exclude that part of the observed changes may relate to issues not directly linked to binding.

To investigate whether pNPY really associates with N-Y2 we have performed a titration experiment, in which up to 10 equivalents of pNPY were added to ^{15}N-labelled N-Y2 (see Fig. 6). The data clearly show concentration-dependent changes of positions of resonances from the N-Y2. Resonances in the segments comprising N-Y2 residues 16-21 and 33-50 are mostly affected. The data point to a low-affinity interaction of pNPY towards N-Y2 without much specificity. To summarize the interaction studies we can say that significant and reliable effects were only detected in the presence of DPC micelles and that all peptides interact with the three N-terminal domains. The interaction with bPP results in the largest changes.

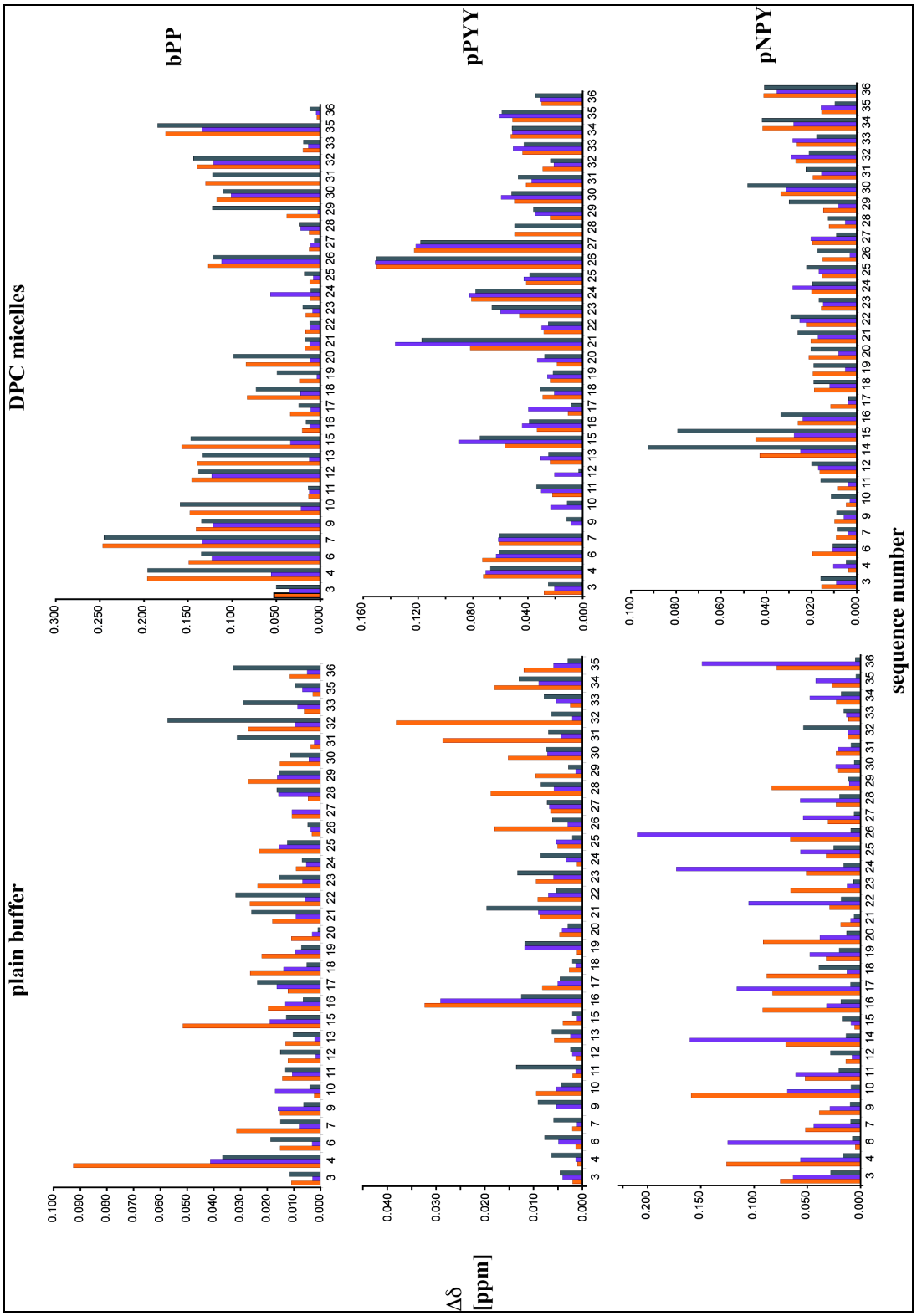


Figure 5: Chemical shift deviation of bPP (top panel), pPYY (middle panel) and pNPY (bottom panel) upon addition of N-Y1(orange bars), N-Y2 (pink bars) and N-Y5 (gray bars) in aqueous buffer (left column) or in the presence of DPC micelles (right column).

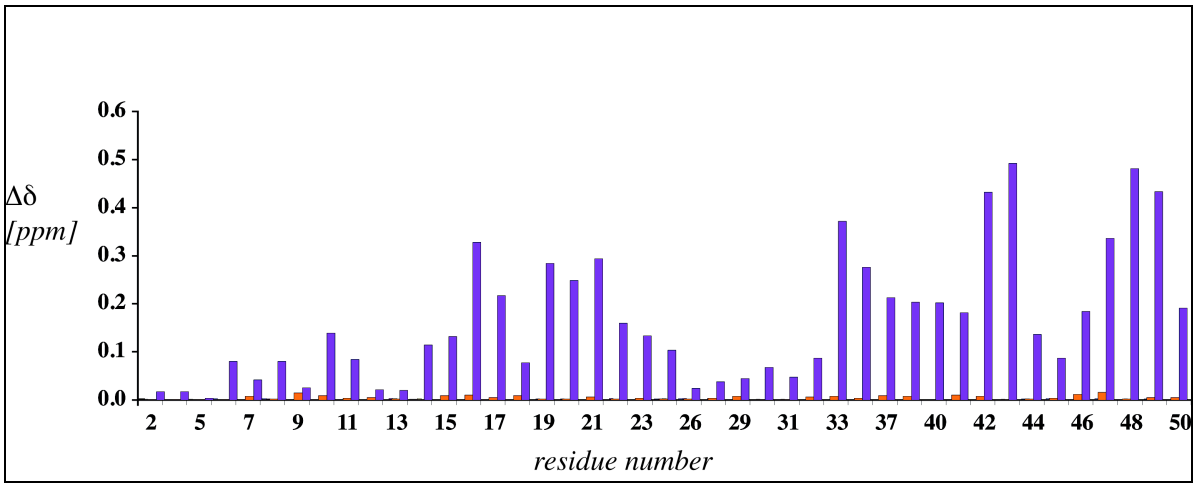


Figure 6: Chemical shift deviation of N-Y2 after addition of 1 and 10 equivalents of pNPY (from left to right). For additional data points at 0.5, 2 and 4 equivalents see Figure S12.

6.2 Discussions

We have postulated that binding of ligands to Y receptors is preceded by association of the ligands to the plasma membrane. Thereby, the apparent concentration of the ligand is increased and the search for the receptor reduced from three to two dimensions [53, 54]. We now studied whether parts of the receptor that protrude into the extracellular compartment may help in transferring ligands, which have accumulated in vicinity of the membrane, into the binding pocket. Such portions of receptor that point into the extracellular space are the N-terminal domains. Herein, we have developed strategies to produce these polypeptides recombinantly in isotopically-enriched form for use in high-resolution NMR studies.

The work has demonstrated that these peptides can all be expressed as soluble fusions to ubiquitin. However, N-Y4 and N-Y1 are degraded in the intracellular milieu, and hence much better yields were obtained using insoluble fusions. Cleavage of the target sequence from the insoluble fusion partner could be obtained by solubilizing the fusion protein in the mild detergent sarcosyl, which proved to be compatible with enzymatic activity of the TEV protease used to cleave the peptide from the fusion protein.

Studies on the structure and dynamics of the peptides using NMR revealed that they are all completely disordered in aqueous buffer. In the presence of phospholipids micelles, segments of most receptor N termini became conformationally stabilized, with the exception of N-Y2, which remained. Otherwise, more (N-Y4) or less stable (N-Y1 or N-Y5) helical segments occurred within the sequences. For all N-terminal peptides chemical shit changes occurred in the presence of DPC micelles, except for N-Y2. This implies that all other peptides associate with the micelle to some extent. Previously, we have made extensive use of the thermodynamic data of Wimley and White for partitioning of single amino acids into the water-membrane interface or the membrane interior[55] to rationalize how peptides interact with phospholipids micelles. A common observation was that the occurrence of the aromatic residues Trp and Tyr help in anchoring peptides in the interface[56]. The partitioning values of the four sequences of the N-terminal domains from the Y receptor subtypes are shown in Figure S7 in the supplementary materials. In N-Y4 a stretch comprising residues 5 to 11 is predicted to show partitioning into the micelle interior. This corresponds exactly to the region that becomes helically structured in the presence of micelles. In case of N-Y2 many negatively charges residues occur throughout the sequence, while they are clustered in the central (unstructured) segment in N-Y4. Even more importantly, many Pro residues are present in N-Y2 that might prevent formation of secondary structure. The sequence of N-Y5 in comparison to N-Y2 is much more amphiphilic in nature, and therefore more likely to favorably interact with the micelles. Again, the regions that become better structured in the presence of DPC micelles correspond to stretches rich in hydrophobic/aromatic residues and hence are predicted to partition into the micelles.

Our interaction studies using chemical shift mapping indicated that bPP strongly interacts with all N-terminal domains, but differences in the sensitivity of certain positions are observed. In contrast, for pPYY the changes are smaller, but more uniform upon addition of the different N termini. For pNPY the changes are very small, although the sequence of pNPY displays more than 80% homology to the one of pPYY. The difference between pNPY and pPYY is very intriguing. We speculate that this difference may be due to structural transitions occurring in pPYY when transiently associating with

the N-terminal domain. pPYY in aqueous buffer displays the so-called PP-fold[57], and tertiary structure is lost when pPYY associates with the membrane surface, or is transferred into a more hydrophobic solvent (e.g. methanol)[58]. Transient binding of PYY to any of the N-terminal domains is expected to alter the equilibrium between membrane-associated peptide, which is devoid of tertiary structure, and the membrane-detached peptide, that could possibly re-adopt its PP fold. Such a change will result in large chemical shift changes and would explain why the differences in pPYY are so much larger than for pNPY upon addition of the different N termini. In addition the changes due to such a structural transition may be much larger than those due to direct contacts, and therefore can possibly explain, why the changes in pPYY are so uniform across the different N-termini. Considering the sensitivity of the chemical shift mapping methodology to structural changes, but on the other hand the inherently low sensitivity of NMR, we consider results from SPR measurements to be more reliable for quantifying (but not for detecting) such interactions. In our Biacore measurements we could detect strongest binding (K_D approx. 50 μM) for the bPP-NY4 interaction, and chemical shift mapping also revealed the largest changes for bPP upon addition of N-Y4.

To summarize this work has described synthetic methods to produce all N-terminal domains in isotopically labeled form in quantities sufficient, for the analysis by various biophysical methods. Structural studies revealed them to be fairly flexible. However, while N-Y2 is fully unfolded residual helical structures were detected in N-Y1 and N-Y5. For the case of N-Y4 we could previously detect a short rather rigid a-helical stretch in the presence of DPC micelles. In contrast to N-Y4, the nascent helical regions of N-Y1 and N-Y5 contain too much residual motion, so that structure calculations did not fully converge towards a-helical structures. All peptides interact with the N terminal domains of N-Y1, N-Y2 and N-Y5, but the interactions are weaker than those previously described for bPP binding to N-Y4.

6.3 Materials and methods

6.3.1 Materials

$^{15}NH_4Cl$ was from Spectra Isotopes (Columbia, USA), d_{38}-DPC- (99%-d), and D_2O was from Cambridge Isotope Laboratories (Andover, Massachusets, USA). 5-doxylstearic acid was from Aldrich (Buchs, Switzerland). Oligonucleotide primers were synthesized by Microsynth GmbH (Balgach, Switzerland).

6.3.2 Expression and purification of N-terminal domains

Depending on their stability against proteolysis the N-terminal domains were either expressed as fusions to ubiquitin (N-Y2 and N-Y5) or to ketosteroidisomerase (N-Y1 and N-Y4). In case of N-Y2 and N-Y5 the amino acid sequence was reverse translated into a DNA sequence taking into account the preferred *E.coli* codon usage including a terminal stop codon and a *Sal*I restriction site. The resulting fragments were purified by electrophoresis and gel extraction and digested with *Sal*I, resulting in fragments that were then blunt-ended on one side and contained *Sal*I-cohesive end on the other end. These fragments were ligated into the pUBK19 vector (gift from T. Kohno, Mitsubishi Kasei Institute of Life Science, Tokyo, Japan), which had been digested with *Nsi*I and *Sal*I and purified before. The resulting plasmids were sequenced and transformed into C41 cells[59].

For production of ^{15}N-labeled peptides M9 minimal media containing ^{15}N-ammoniumchloride as the sole nitrogen source was used, otherwise expression was done on LB medium. In each case 1 liter of medium containing 50 mg/ml kanamycin was inoculated with 10 ml of an overnight LB culture. Cultures were induced at OD_{600} around 0.5 with 0.4 mM IPTG. LB- and minimal medium cultures were grown under induction for 4 h and 11 h, respectively. Cells were harvested by centrifugation on a Sorval GSA rotor at 4 °C and stored at -20 °C. The cell pellets were thawed on ice for 1 h and resuspended in 25 ml denaturing basic buffer (50 mM Tris pH 8, 6 M GdmHCl, 100 mM NaCl, 1 mM β-mercaptoethanol). The suspension was lysed by sonication on ice. The ubiquitin fusion proteins were purified by Ni-NTA chromatography. Refolding was achieved by applying a linear gradient to exchange the denaturing basic buffer to native

binding buffer (50 mM Tris pH 8, 100 mM NaCl, 1 mM β-mercaptoethanol, 20 mM imidazole), and the protein was eluted with binding buffer containing 200 mM imidazole. The eluates were diluted 10-fold with basic buffer (50 mM Tris pH 8, 100 mM NaCl, 1 mM β-mercaptoethanol) and a 1 mg/ml YUH-solution (for expression and purification of YUH see supplementary S8) was added in a 20-fold dilution. The cleavage reactions were allowed to proceed for 3 hours at 37 °C.

In case of N-Y1 and N-Y4 the DNA sequences were subcloned from wt cDNA of the corresponding Y receptor (University of Missouri-Rolla (UMR) cDNA Resource Center by PCR). During PCR, a GSGSGS linker followed by TEV cleavage sequence was introduced at the N terminus of the target sequence. After digestion with XhoI and EspI, the fragments were ligated with T4 DNA ligase into the pET31b vector, which had been digested with XhoI and EspI. The correctness of the constructs was verified by DNA sequencing (Synergene Biotech, Switzerland). The resulting plasmids were transformed into BL21(DE3) for expression. For production of ^{15}N-labeled peptides M9 minimal media containing ^{15}N-ammoniumchloride as the sole nitrogen source was used, otherwise expression was done in LB medium. In each case 1 liter of medium containing 50 mg/ml kanamycin was inoculated with 10 ml of an overnight LB culture. Cultures were induced at OD_{600} of 0.7 with 1 mM IPTG, harvested after 5 hours by centrifugation on a Sorval GSA rotor at 4 °C and the pellets were stored at -20 °C.

The fusion proteins were purified from inclusion bodies by Ni-NTA chromatography in presence of 6 M GdmHCl. After removal of GdmHCl by dialysis the precipitated fusion protein was solubilized in 50 mM Tris pH 8.0 in the presence of 2% N-lauryl sarcosine upon sonication to a final concentration of 2 mg/ml. The resulting solution was dialyzed against a 20-fold excess of 50 mM Tris, pH 8.0 for 4-6 times. The solution was diluted 10 times with 50 mM Tris pH 8.0 and EDTA and DTT were added to a final concentration of 0.5 mM and 1 mM, respectively. TEV protease (for expression and purification of TEV protease see supplementary materials S9) was added to a final concentration of 100 μM and the cleavage mixture was incubated at 4°C over night.

All target peptides were finally purified by C18-RP-HPLC (Vydac, USA) by using a water/acetonitrile/0.1% TFA gradient. Yields ranged from 3 mg to 20 mg peptide from 1 liter of culture. The mass of all peptides was confirmed by MALDI-TOF MS or ESI MS: N-Y1: 4532.9 Da (theoretical value: 4533.0 Da); ^{15}N-N-Y1: 4587.0 Da (theoretical value: 4587.0 Da); N-Y2: 5509.3 Da (theoretical value: 5510.0 Da); ^{15}N-N-Y2: 5568.0 Da (theoretical value: 5570.0 Da); N-Y4: 4554.0 Da (theoretical value: 4556.1 Da); ^{15}N-N-Y4: 4614.0 Da (theoretical value: 4611.1 Da); N-Y5: 6053.7 (theoretical value: 6053.4); ^{15}N-N-Y5: 6119.5 Da (theoretical value: 6118.4 Da).

6.3.3 NMR spectroscopy

For studies of structure or backbone dynamics 1 mM solution of the peptides at pH 5.6, 20 mM d_{13}-MES, 300 mM d_{38}-DPC were used. All spectra were recorded on an AV-700 Bruker NMR spectrometer at 310 K. Chemical shifts were calibrated to the water line at 4.63 ppm and nitrogen shifts were referenced indirectly to liquid NH_3. The spectra were processed using the Bruker Topspin2.0 software and transferred into the XEASY[60] or CARA[61] programs for further analysis.

For chemical shift assignments 3D ^{15}N-resolved TOCSY and NOESY[62] were used. In case of N-Y5 we decided to use ^{13}C,^{15}N labeling in combination with experiments that directly correlate sequential amide moieties[63]. In general experiments used coherence selection schemes via pulsed-field gradients[64] and sensitivity-enhancement building blocks[65] whenever possible. Upper-distance limits for structure calculations of N-Y1 were derived from a 70 ms NOESY spectrum[66]. Structures were calculated in the program CYANA[67], that uses restraint molecular dynamics in torsion angle space, and the implemented standard simulated annealing protocol in CYANA was used for that task. A proton-detected version of the steady-state ^{15}N[68] heteronuclear Overhauser effect sequence were used for measurement of the hetereonuclear NOE[69]. Therein, the buildup of the NOE was achieved through a pulse train of 120 degree proton pulses separated by 5 ms over a period of 3 seconds.

For measurements of interactions by chemical shift mapping methodology 0.1 mM solutions of the ^{15}N-labeled neurohormones were mixed with the corresponding peptides from the N-terminal domains, and the deviations of peak positions were extracted from the [^{15}N,^{1}H]-HSQC spectra and computed according to $\Delta\delta=SQRT(\Delta(^{1}H)^{2}+0.2 \text{ X } \Delta(^{15}N)^{2})$. Particular care was taken to ensure that no shifts in pH occurred when adding the N-Y peptides. In case of addition of various equivalents of pNPY to ^{15}N-labelled N-Y2 in the presence of DPC micelles (please describe the conditions here)

6.4 Acknowledgments

We would like to thank for financial support from the Swiss national science foundation (grant No. 3100A0-11173).

6.5 References

1. Ma, P. and R. Zemmel, *Value of novelty?* Nat Rev Drug Discov, 2002. 1(8): p. 571-2.
2. Scheerer, P., et al., *Crystal structure of opsin in its G-protein-interacting conformation.* Nature, 2008. 455(7212): p. 497-502.
3. Park, J.H., et al., *Crystal structure of the ligand-free G-protein-coupled receptor opsin.* Nature, 2008. 454(7201): p. 183-7.
4. Hieke, M., H. Zettl, and M. Schubert-Zsilavecz, *[GPCR--structure of the human beta-2 receptor clarified].* Pharm Unserer Zeit, 2008. 37(3): p. 191-2.
5. Hanson, M.A., et al., *A specific cholesterol binding site is established by the 2.8 A structure of the human beta2-adrenergic receptor.* Structure, 2008. 16(6): p. 897-905.
6. Warne, T., et al., *Structure of a beta1-adrenergic G-protein-coupled receptor.* Nature, 2008. 454(7203): p. 486-91.
7. Rosenbaum, D.M., et al., *GPCR engineering yields high-resolution structural insights into beta2-adrenergic receptor function.* Science, 2007. 318(5854): p. 1266-73.
8. Jaakola, V.P., et al., *The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist.* Science, 2008. 322(5905): p. 1211-7.
9. Murakami, M. and T. Kouyama, *Crystal structure of squid rhodopsin.* Nature, 2008. 453(7193): p. 363-7.
10. Kobilka, B.K., *G protein coupled receptor structure and activation.* Biochim Biophys Acta, 2007. 1768(4): p. 794-807.
11. Harmar, A.J., *Family-B G-protein-coupled receptors.* Genome Biol, 2001. 2(12): p. REVIEWS3013.
12. O'Hara, P.J., et al., *The ligand-binding domain in metabotropic glutamate*

receptors is related to bacterial periplasmic binding proteins. Neuron, 1993. 11(1): p. 41-52.
13. Okamoto, T., et al., *Expression and purification of the extracellular ligand binding region of metabotropic glutamate receptor subtype 1.* J Biol Chem, 1998. 273(21): p. 13089-96.
14. Trumpp-Kallmeyer, S., et al., *Towards understanding the role of the first extracellular loop for the binding of peptide hormones to G-protein coupled receptors.* Pharm Acta Helv, 1995. 70(3): p. 255-62.
15. DeMartino, J.A., et al., *The amino terminus of the human C5a receptor is required for high affinity C5a binding and for receptor activation by C5a but not C5a analogs.* J Biol Chem, 1994. 269(20): p. 14446-50.
16. Dettin, M., et al., *CCR5 N-terminus peptides enhance X4 HIV-1 infection by CXCR4 up-regulation.* Biochem Biophys Res Commun, 2003. 307(3): p. 640-6.
17. Ho, H.H., D. Du, and M.C. Gershengorn, *The N terminus of Kaposi's sarcoma-associated herpesvirus G protein-coupled receptor is necessary for high affinity chemokine binding but not for constitutive activity.* J Biol Chem, 1999. 274(44): p. 31327-32.
18. Blanpain, C., et al., *Multiple charged and aromatic residues in CCR5 amino-terminal domain are involved in high affinity binding of both chemokines and HIV-1 Env protein.* J Biol Chem, 1999. 274(49): p. 34719-27.
19. Hawtin, S.R., et al., *The N-terminal juxtamembrane segment of the V1a vasopressin receptor provides two independent epitopes required for high-affinity agonist binding and signaling.* Mol Endocrinol, 2005. 19(11): p. 2871-81.
20. Koller, D., et al., *The extreme N-terminus of the calcitonin-like receptor contributes to the selective interaction with adrenomedullin or calcitonin gene-related peptide.* FEBS Lett, 2002. 531(3): p. 464-8.
21. Monteclaro, F.S. and I.F. Charo, *The amino-terminal domain of CCR2 is both necessary and sufficient for high affinity binding of monocyte chemoattractant protein 1. Receptor activation by a pseudo-tethered ligand.* J Biol Chem, 1997. 272(37): p. 23186-90.
22. Chen, Y., et al., *The amino terminus and the third extracellular loop of CX3CR1 contain determinants critical for distinct receptor functions.* Mol Pharmacol, 2006. 69(3): p. 857-65.
23. Pervushin, K.V., et al., *Three-dimensional structure of (1-71)bacterioopsin solubilized in methanol/chloroform and SDS micelles determined by ^{15}N-^{1}H heteronuclear NMR spectroscopy.* Eur. J. Biochem., 1994. 219(1-2): p. 571-83.
24. Ulfers, A.L., A. Piserchio, and D.F. Mierke, *Extracellular domains of the neurokinin-1 receptor: structural characterization and interactions with substance P.* Biopolymers, 2002. 66(5): p. 339-49.
25. Grace, C.R., et al., *Structure of the N-terminal domain of a type B1 G protein-coupled receptor in complex with a peptide ligand.* Proc Natl Acad Sci U S A, 2007. 104(12): p. 4858-63.
26. Yeagle, P.L., et al., *Structures of the intradiskal loops and amino terminus of the G-protein receptor, rhodopsin.* J Pept Res, 2000. 55(6): p. 455-65.
27. Chopra, A., et al., *Solution structure of the sixth transmembrane helix of the G-protein-coupled receptor, rhodopsin.* Biochim Biophys Acta, 2000. 1463(1): p. 1-

5.
28. Cherezov, V., et al., *High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor.* Science, 2007. 318(5854): p. 1258-65.
29. Rasmussen, S.G.F., et al., *Crystal structure of the human beta(2) adrenergic G-protein-coupled receptor.* Nature, 2007. 450: p. 383-U4.
30. Michel, M.C., et al., *XVI. International Union of Pharmacology recommendations for the nomenclature of neuropeptide Y, peptide YY, and pancreatic polypeptide receptors.* Pharmacol Rev, 1998. 50(1): p. 143-50.
31. Larhammar, D., et al., *Origins of the many NPY-family receptors in mammals.* Peptides, 2001. 22(3): p. 295-307.
32. Berglund, M.M., et al., *Studies of the human, rat, and guinea pig Y4 receptors using neuropeptide Y analogues and two distinct radioligands.* Peptides, 2001. 22(3): p. 351-6.
33. Fields, G.B. and S.P. Colowick, *Solid-Phase Peptide Synthesis*. Methods Enzymol. 1997, London: Academic Press.
34. Merrifield, B., *Concept and early development of solid-phase peptide synthesis.* Methods Enzymol., 1997. 289: p. 3-13.
35. Sahdev, S., S. Khattar, and K. Saini, *Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies.* Mol Cell Biochem, 2008. 307(1-2): p. 249-264.
36. LaVallie, E.R., et al., *Enzymatic and chemical cleavage of fusion proteins.* Curr. Protoc. Mol. Biol., 2001. 16: p. 4.9-4.17
37. KULIOPULOS, A. and C. WALSH, *PRODUCTION, PURIFICATION, AND CLEAVAGE OF TANDEM REPEATS OF RECOMBINANT PEPTIDES.* J Am Chem Soc, 1994. 116(11): p. 4599-4607.
38. Bornstein, P. and G. Balian, *Cleavage at Asn-Gly bonds with hydroxlyamine.* Meth. Enzymol. , 1977. 47: p. 132-145.
39. Kohno, T., et al., *A new general method for the biosynthesis of stable isotope-enriched peptides using a decahistidine-tagged ubiquitin fusion system: an application to the production of mastoparan-X uniformly enriched with ^{15}N and $^{15}N/^{13}C$.* J. Biomol. NMR, 1998. 12(1): p. 109-21.
40. Kapust, R.B., et al., *The P1' specificity of tobacco etch virus protease.* Biochem Biophys Res Commun, 2002. 294(5): p. 949-55.
41. Kapust, R.B., et al., *Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency.* Protein Eng, 2001. 14(12): p. 993-1000.
42. Mohanty, A., C. Simmons, and M. Wiener, *Inhibition of tobacco etch virus protease activity by detergents.* Protein Expres Purif, 2003. 27(1): p. 109-114-PII S1046-5928(02)00589-2.
43. Wüthrich, K., *NMR of Proteins and Nucleic Acids*. 1986: Wiley. Wiley, New York.
44. Wishart, D., B. Sykes, and F. Richards, *Relationship between nuclear magnetic resonance chemical shift and protein secondary structure.* J Mol Biol, 1991. 222(2): p. 311-33.
45. Palmer, A.G., *NMR Probes of molecular dynamics: Overview and comparison*

with other techniques. Ann. Rev. Biophys. Biomol. Struct., 2001. 30: p. 129-155.
46. Brown, L.R., C. Bösch, and K. Wüthrich, *Location and orientation relative to the micelle surface for glucagon in mixed micelles with dodecylphosphocholine: EPR and NMR studies.* Biochim Biophys Acta, 1981. 642(2): p. 296-312.
47. Zou, C., et al., *Studies of the structure of the N-terminal domain from the Y4 receptor, a G-protein coupled receptor, and its interaction with hormones from the NPY family.* ChemBioChem, 2008. 9: p. 2276-2284.
48. Wüthrich, K., M. Billeter, and W. Braun, *Polypeptide Secondary Structure Determination by Nuclear Magnetic Resonance Observation of Short Proton-Proton Distances.* J. Mol. Biol., 1984. 180: p. 715-740.
49. Zerbe, O., et al., *Recognition of neurohormones of the NPY family by their receptors.* Journal of receptor and signal transduction research, 2006. 26(5-6): p. 487-504.
50. Bader, R., et al., *Structure and Dynamics of Micelle-bound Neuropeptide Y: Comparison with unligated NPY and Implications for Receptor Selection.* J. Mol. Biol., 2001. 305: p. 307-392.
51. Bettio, A., M.C. Dinger, and A.G. Beck-Sickinger, *The neuropeptide Y monomer in solution is not folded in the pancreatic- polypeptide fold.* Protein Sci, 2002. 11(7): p. 1834-44.
52. Cowley, D.J., et al., *Structure of neuropeptide Y dimer in solution.* Eur. J. Biochem., 1992. 205(3): p. 1099-106.
53. Sargent, D.F. and R. Schwyzer, *Membrane lipid phase as catalyst for peptide-receptor interactions.* Proc. Natl Acad. Sci. U S A, 1986. 83(16): p. 5774-8.
54. Schwyzer, R., *Membrane-assisted molecular mechanism of neurokinin receptor subtype selection.* EMBO Journal EMBO J, 1987. 6(8): p. 2255-9.
55. Wimley, W.C., T.P. Creamer, and S.H. White, *Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides.* Biochemistry, 1996. 35(16): p. 5109-24.
56. Ridder, A.N., et al., *Analysis of the role of interfacial tryptophan residues in controlling the topology of membrane proteins.* Biochemistry, 2000. 39(21): p. 6521-8.
57. Lerch, M., M. Mayrhofer, and O. Zerbe, *Structural similarities of micelle-bound peptide YY (PYY) and neuropeptide Y (NPY) are related to their affinity profiles at the Y receptors.* J. Mol. Biol., 2004. 339(5): p. 1153-68.
58. Neumoin, A., et al., *Probing the Formation of Stable Tertiary Structure in a Model Miniprotein at Atomic Resolution: Determinants of Stability of a Helical Hairpin.* Journal of the American Chemical Society, 2007: p. .
59. Miroux, B. and J.E. Walker, *Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels.* J. Mol. Biol., 1996. 260(3): p. 289-98.
60. Bartels, C., et al., *The program XEASY for computer-supported spectral analysis of biological macromolecules.* J. Biomol. NMR, 1995. 6: p. 1-10.
61. Keller, R., *The Computer Aided Resonance Assignment*. 2004, Goldau: CANTINA Verlag.
62. Marion, D., et al., *Overcoming the Overlap Problem in the Assignment of ^{1H} NMR Spectra of Larger Proteins by Use of Three-Dimensional Heteronuclear 1H-15N*

Hartmann-Hahn Multiple-Quantum Coherence and Nuclear Overhauser-Multiple Quantum Coherence Spectroscopy: Application to Interleukin 1b. Biochemistry, 1989. 28: p. 6150-6156.
63. Weisemann, R., H. Ruterjans, and W. Bermel, *3D triple-resonance NMR techniques for the sequential assignment of NH and 15N resonances in 15N- and 13C-labelled proteins.* J Biomol NMR, 1993. 3(1): p. 113-20.
64. Keeler, J., et al., *Pulsed-field gradients: theory and practice.* Methods Enzymol, 1994. 239: p. 145-207.
65. Kay, L.E., P. Keifer, and T. Saarien, *Pure absorption gradient enhanced heteronuclear single-quantum correlation sepctroscopy with improved sensitivity.* J. Am. Chem. Soc., 1992. 114: p. 10663-10665.
66. Kumar, A., R.R. Ernst, and K. Wüthrich, *A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules.* Biochem. Biophys. Res. Commun., 1980. 95: p. 1-6.
67. Guntert, P., *Automated NMR structure calculation with CYANA.* Methods Mol Biol, 2004. 278: p. 353-78.
68. Dmitriev, O., et al., *Structure of the membrane domain of subunit b of the Escherichia coli F0F1 ATP synthase.* J Biol Chem, 1999. 274(22): p. 15598-604.
69. Noggle, J.H. and R.E. Schirmer, *The Nuclear Overhauser Effect - Chemical Applications*. 1971, New York: Academic Press.

6.6 Supplementary materials

In tables S1-S4 chemical shifts were referenced to the water line taken at 4.63 ppm at 310K. The ^{15}N scale was derived indirectly by multiplying the frequency of 0 ppm for protons (the Bruker parameter SF) by 0.101329118. Chemical shifts have been deposited in the BMRB data base under deposition codes 80.8933262 (N-Y1), 80.6873033 (N-Y2) and 80.74817093 (N-Y5).

Table S1: Chemical shifts of N-Y1 in the presence of DPC micelles.

	H^N	H'^{α}	H^{β}	others
Met 1	-	-	-, -	γCH_2 -, -; εCH_3 -
Asn 2	-	-	-, -	δNH_2 -, -
Ser 3	8.52	4.42	-, -	γOH -
Thr 4	8.32	4.28	4.01	γCH_3 1.17; γOH -
Leu 5	8.23	4.03	1.86, 1.86	γH 1.17; δCH_3 0.79, 0.72
Phe 6	7.97	4.46	2.96, 3.18	δH 7.19; εH -, - ; ζH -
Ser 7	8.02	4.31	3.86, 3.86	γOH -
Gln 8	8.25	4.28	2.04, 2.04	γCH_2 2.33; εNH_2 7.44
Val 9	7.90	3.94	2.05	γCH_3 0.88
Glu 10	8.26	4.09	1.85, 1.88	γCH_2 2.33; εH -
Asn 11	8.29	4.51	2.69, 2.69	δNH_2 7.58, 6.86
His 12	8.32	4.57	3.05, 3.19	δ^1NH -; δ^2H 7.09; ε^1H -; ε^2NH -
Ser 13	8.18	4.40	3.79, 3.79	γOH -
Val 14	8.15	4.00	2.01	γCH_3 0.81
His 15	8.30	4.61	3.09, 3.09	δ^1NH -; δ^2H 7.05; ε^1H -; ε^2NH -
Ser 16	8.22	4.62	3.77, 3.80	γOH -
Asn 17	8.17	4.19	2.63, 2.63	δNH_2 7.50, 6.81
Phe 18	8.22	4.50	2.99, 3.10	δH 7.17; εH 7.25; ζH -
Ser 19	8.17	4.29	3.79, 3.79	γOH -
Glu 20	8.38	4.14	1.96, 1.96	γCH_2 2.22; εH -
Lys 21	8.15	4.18	1.71, 1.71	γCH_2 1.37; δCH_2 1.94, 2.02; εCH_2 2.74; ζNH_3^+ -
Asn 22	8.13	4.57	2.68, 2.68	δNH_2 7.49, 6.81
Ala 23	8.13	4.11	1.36	
Gln 24	8.13	4.15	2.02, 2.02	γCH_2 2.31; εNH_2 7.49
Leu 25	8.01	4.16	1.64, 1.64	γH 1.54; δCH_3 0.88, 0.81
Leu 26	7.83	4.17	1.57, 1.57	γH 1.46; δCH_3 0.83, 0.79
Ala 27	7.74	4.19	1.23	
Phe 28	7.97	4.55	3.13, 3.13	δH 7.20; εH 7.04 ; ζH -
Glu 29	8.33	4.20	1.87, 1.87	γCH_2 2.18; εH -
Asn 30	8.30	4.62	2.78, 2.78	δNH_2 7.58, 6.86
Asp 31	8.19	4.50	2.58, 2.58	δH -
Asp 32	8.07	4.29	2.78, 2.78	δH -
Cys 33	8.29	4.75	2.59, 2.65	γSH -
His 34	8.36	4.66	3.09, 3.09	δ^1NH -; δ^2H 7.13; ε^1H -; ε^2NH -
Leu 35	8.32	4.51	1.62, 1.62	γH 1.44; δCH_3 0.86
Pro 36		4.44	1.91, 1.96	γCH_2 2.19,2.22 ; δCH_2 3.78, 3.83
Leu 37	8.05	4.18	1.57, 1.57	γH -; δCH_3 0.87, 0.81
Ala 38	8.18	4.27	1.32	
Met 39	8.22	4.38	2.06	γCH_2 2.55, 2.66; εCH_3 1.96
Ile 40	7.38	4.01	1.79	γCH_2 1.09, 1.37; δCH_3 0.82

Table S2: Amide proton and ^{15}N chemical shifts of N-Y1.

	N	HN
Met 1		
Asn 2		
Ser 3	116.26	8.52
Thr 4	116.44	8.36
Leu 5	122.7	8.27
Phe 6	115.8	7.99
Ser 7	114.84	8.04
Gln 8	120.6	8.25
Val 9	118.8	7.9
Glu 10	122.35	8.27
Asn 11	118.23	8.29
His 12	119.81	8.26
Ser 13	115.76	8.24
Val 14	120.68	8.18
His 15	120.62	8.38
Ser 16	116.22	8.2
Asn 17	119.23	8.2
Phe 18		
Ser 19	115.9	8.21
Glu 20	122.3	8.41
Lys 21	120.03	8.16
Asn 22	118.11	8.13
Ala 23	123.09	8.16
Gln 24	117.35	8.13
Leu 25	120.96	8.03
Leu 26	118.6	7.84
Ala 27	122.05	7.74
Phe 28	117.83	7.98
Glu 29	120.35	8.34
Asn 30	118.61	8.31
Asp 31	120.03	8.31
Asp 32	118.07	8.08
Cys 33	120.08	8.47
His 34	119.74	8.39
Leu 35	122.68	8.39
Pro 36		
Leu 37	120.63	8.05
Ala 38	122.21	8.2
Met 39	118.15	8.23
Ile 40	122.86	7.38

Table S3: Amide proton and ^{15}N chemical shifts of N-Y2.

	N	HN
Met 1		
Gly 2	112.1	8.59
Pro 3		
Ile 4	120.1	8.16
Gly 5	112.9	8.42
Ala 6	123.6	8.07
Glu 7	119.4	8.39
Ala 8	123.9	8.14
Asp 9	119.1	8.12
Glu 10	120.9	8.3
Asn 11	118.7	8.37
Gln 12	120.3	8.15
Thr 13	115.7	8.15
Val 14	122.2	8.09
Glu 15	123.8	8.33
Glu 16	121.5	8.26
Met 17	121.1	8.22
Lys 18	122.8	8.18
Val 19	121	8.02
Glu 20	123.9	8.37
Gln 21	120.6	8.15
Tyr 22	120.4	8.12
Gly 23	109.8	8.06
Pro 24		
Gln 25	119.8	8.49
Thr 26	114.7	8.08
Thr 27	118.5	8.09
Pro 28		
Arg 29	121	8.33
Gly 30	109.4	8.28
Glu 31	119.9	8.15
Leu 32	122.8	8.19
Val 33	122.6	8.04
Pro 34		
Asp 35	121.3	8.23
Pro 36		
Glu 37	121.8	8.3
Pro 38		
Glu 39	120.1	8.37
Leu 40	123.3	8.18
Ile 41	121.2	7.98
Asp 42	123.9	8.26
Ser 43	117.8	8.36
Thr 44	114.9	8.21

Lys 45	122.2	7.83
Leu 46	122.2	7.92
Ile 47	120.2	7.83
Glu 48	124.4	8.26
Val 49	121.4	8.06
Gln 50	128.5	7.89

Table S4: Amide proton and ^{15}N chemical shifts of N-Y4.

Met 1		
Asn 2	8.171	119.838
Thr 3	8.29	114.787
Ser 4	8.26	127.348
His 5		
Leu 6		
Leu 7	8.006	121.818
Ala 8	7.953	123.247
Leu 9	7.817	119.79
Leu 10	7.915	121.832
Leu 11	7.897	123.675
Pro 12		
Lys 13	8.257	120.984
Ser 14	8.221	117.544
Pro 15		
Gln 16	8.338	119.671
Gly 17	8.243	109.457
Glu 18	8.225	120.207
Asn 19	8.454	119.731
Arg 20	8.252	121.321
Ser 21	8.211	116.37
Lys 22	8.092	123.461
Pro 23		
Leu 24	8.423	122.478
Gly 25	8.315	109.153
Thr 26	7.874	115.514
Pro 27		
Tyr 28		
Asn 29	8.083	120.162
Phe 30	7.989	120.843
Ser 31	8.069	116.045
Glu 32	8.133	121.685
His 33	8.288	117.461
Cys 34		
Gln 35	8.43	121.02
Asp 36	8.224	120.984
Ser 37	8.096	115.276
Val 38	8.002	120.362

Asp 39	8.251	122.983
Val 40	7.918	119.379
Met 41	7.85	128.423

Table S5: Amide proton and ^{15}N chemical shifts of N-Y5.

	N	HN
Met 1		
Ser 2		
Phe 3	118.81	7.84
Tyr 4	118.03	7.63
Ser 5	115.85	8.1
Lys 6	122.27	8.25
Gln 7	119.04	8.18
Asp 8	119.37	8.03
Tyr 9	120.84	8.43
Asn 10	119.6	8.25
Met 11	120.23	8.22
Asp 12	120.96	8.38
Leu 13	121.09	7.9
Glu 14	120.56	8.39
Leu 15	121.51	8.39
Asp 16	118.01	8.47
Glu 17	118.35	8.04
Tyr 18	117.72	7.86
Tyr 19	118.5	8.1
Asn 20	117.52	8.21
Lys 21	118.77	7.9
Thr 22	113.27	7.8
Leu 23	121.69	7.83
Ala 24	122.63	7.76
Thr 25	112.49	7.92
Glu 26	122.74	8.19
Asn 27	120.23	8.43
Asn 28	119.29	8.35
Thr 29	114.05	8.05
Ala 30	125.61	8.17
Ala 31	122.26	8.07
Thr 32	112.43	7.92
Arg 33	122.34	8.32
Asn 34	119.25	8.36
Ser 35	115.51	8.17
Asp 36	121.69	8.26
Phe 37	120	8.11
Pro 38		
Val 39	117.15	8.04
Trp 40	122.72	8.12

Asp 41	118.58	8.12
Asp 42	118.49	7.95
Tyr 43	119.01	7.82
Lys 44	121.02	7.75
Ser 45	115.33	8.12
Ser 46	117.18	8.06
Val 47	118.82	7.84
Asp 48	122.02	8.04
Asp 49	119.12	8.03
Leu 50	121.18	7.9
Gln 51	124.45	7.67

Figure S6: Expansion of the 70 ms NOESY of N-Y1, displaying the assignment of residues 24 to 31.

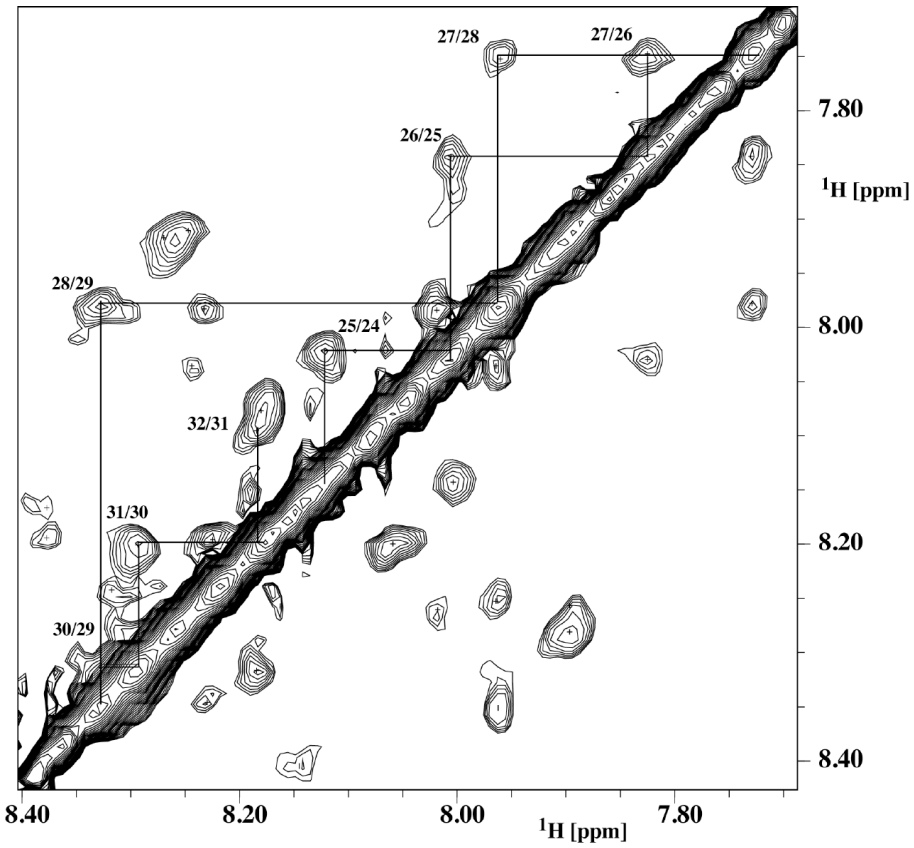
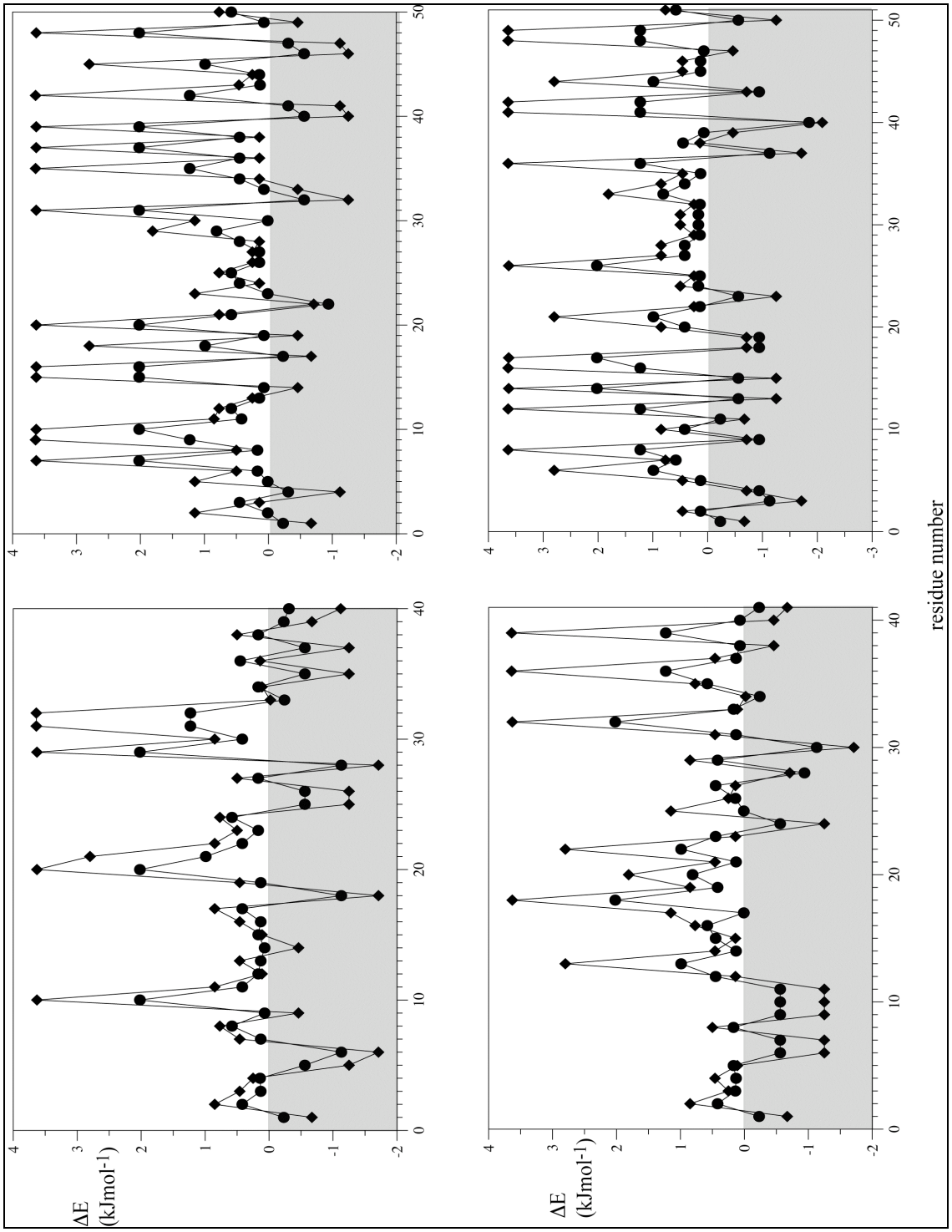


Figure S7: Free energies for partitioning residues of the Y receptor N-terminals (top left: N-Y1, top right: N-Y2, bottom left: N-Y4 and bottom right: N-Y5) domains into the water-membrane interface (circles) or into the membrane interior (diamonds) (data taken from ref. 24). Regions of favorable values are shaded in grey.



S8: Expression and purification of YUH

Yeast ubiquitin hydrolase (YUH) was expressed with a C-terminal hexahistidine tag and purified on a Ni-NTA column. The plasmid coding for the YUH-construct pYUHK20b was a generous gift from Toshiyuki Kohno (Mitsubishi Kasei Institute of Life Science, Tokyo, Japan). 5 ml of LB-broth containing 50 μg/ml kanamycin were inoculated with a colony of BL21 DE3 pYUHK20b cells, streaked onto plate from a glycerol stock, and incubated for 12 h at 37 ˚C and 220 rpm. 0.5 liter of LB-broth containing 50 μg/ml kanamycin were inoculated with the 5 ml overnight culture and incubated at 37 ˚C and 240 rpm in a 2 l Erlenmeyer flask. The culture was induced with 0.4 mM IPTG at an OD_{600} of 0.6-0.7 and grown for another 5 h to a final OD_{600} of around 5. The culture was harvested by centrifugation at 5000 rpm and 4 ˚C for 20 min. 7 g of wet biomass were obtained from 1 l of culture. The cell pellet was frozen at -20 ˚C.

The cell pellet was thawed on ice and resuspended in 40 ml resuspension buffer (50 mM Tris pH 8, 100 mM NaCl, 1 mM β-mercaptoethanol). 9 mg of lysozyme were added and the mixture incubated on ice for 15 min. The resuspension mixture was sonicated on ice with a Branson Digital Sonifier. The lysate was centrifuged twice at 190000 rpm and 4 ˚C for 45 min and loaded onto a 10 ml column volume (CV) Ni-NTA-agarose column previously equilibrated with running buffer (50 mM Tris pH 8, 100 mM NaCl, 1 mM β-mercaptoethanol, 10 mM imidazole). Bound protein was eluted with elution buffer (50 mM Tris pH 8, 100 mM NaCl, 1 mM β-mercaptoethanol, 100 mM imidazole). The eluate was confirmed to contain the target protein by SDS-PAGE. To 10 ml of eluate 1.1 ml of glycerol were added to yield a final glycerol concentration of 10%. The YUH-concentration of this mixture was determined by a Bradford assay to be 7 mg/ml. This solution was stored at -20 °C in 1 ml aliquots.

S9: Expression and purification of TEV protease

The plasmid pTH24[1] was transformed to Rosetta(DE3) pLys cells. 0.5 ml of an overnight LB culture containing 100 μg/ml ampicillin and 34 μg/ml chloramphenicol were used to inoculate 1 liter of LB also containing 100 μg/ml ampicillin and 34 μg/ml chloramphenicol. The culture was incubated at 37 °C. When it reached an OD_{600} of 0.6 it

was induced with 0.1 mM IPTG and the temperature was lowered to 20 °C. After 20 hours the cells were harvested by centrifugation and the cell pellet was stored at –20 °C.

The cell pellet from 1 liter culture was resuspended in 40 ml washing buffer (50 mM sodium phosphate, 300 mM NaCl, 20 mM imidazole, pH 7). The resuspension mixture was sonicated on ice with a Branson Digital Sonifier. After centrifugation at 4 °C at 30000 g for 30 min the supernatant was loaded onto a Ni-NTA column. Unbound protein was eluted with washing buffer and bound protein was eluted with elution buffer (washing buffer with 200 mM imidazole). EDTA and DTT were added to a final concentration of 2 and 10 mM, respectively. 10 ml eluate were dialyzed over night against 1 liter dialysis buffer (25 mM sodium phosphate, 200 mM NaCl, 2 mM EDTA, 2 mM DTT, pH 8) at 4 °C. 10% glycerol were added and the protease solution was stored at –20 °C. Susanne van den Berg, Per-Ake Lofdahl, Torleif Hard, Helena Berglund, Improved solubility of TEV protease by directed evolution, 2006, J. Biotechnol., 121, 291-298

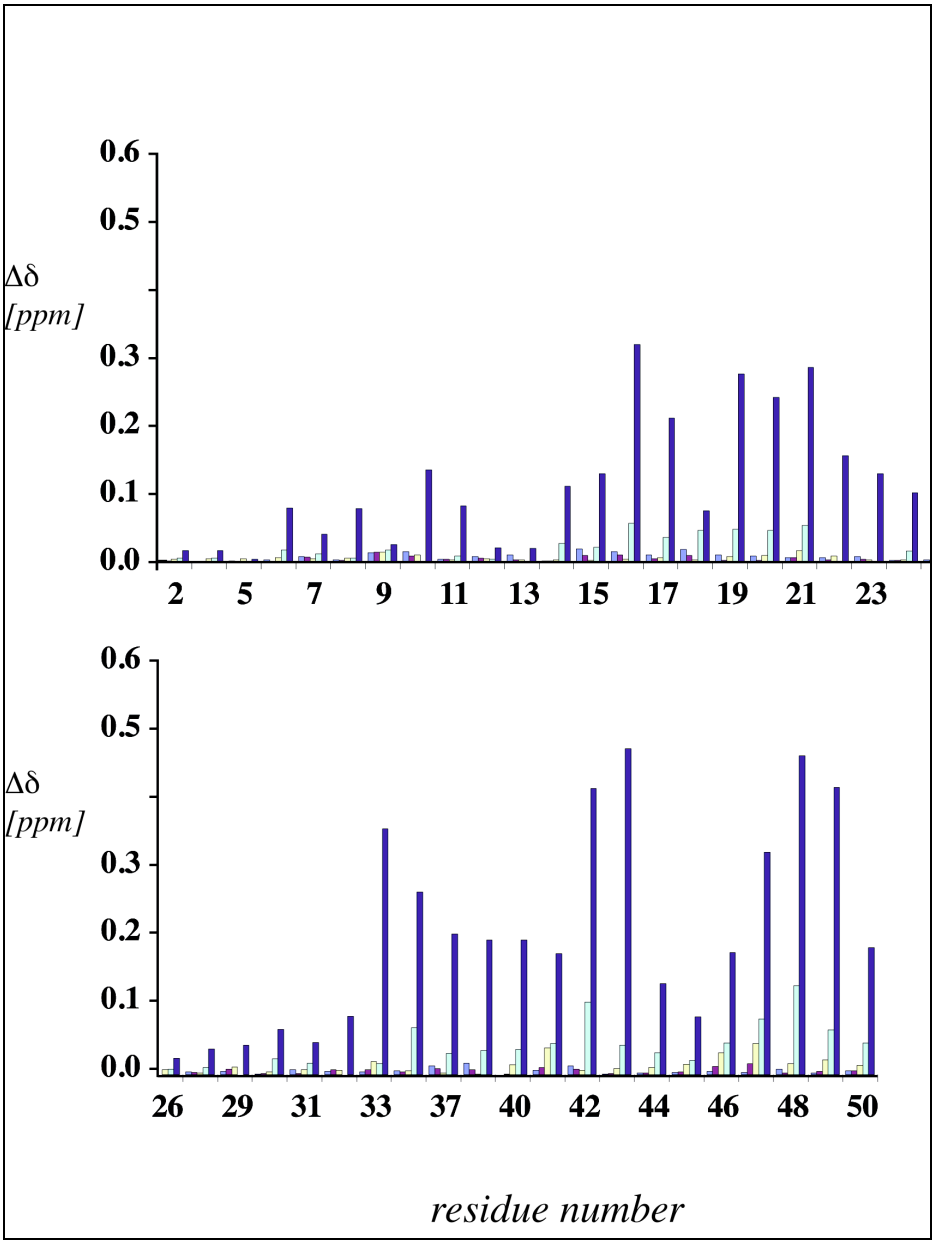
Table S10: MS fingerprinting of cleavage products from the N-Y4 expression:

Exp.	Theor.	_Mass	Peptide	fragment
2701	2701.279	0.278	(L)PKSPQGENRSKPLGTPYNFSEHCQ(D)	12-35
2701	2701.482	0.481	MNTSHLLALLLPKSPQGENRSKPLG(T)	1-25
2899	2899.582	0.582	MNTSHLLALLLPKSPQGENRSKPLGTP(Y)	1-27
3176	3176.688	0.688	MNTSHLLALLLPKSPQGENRSKPLGTPYN(F)	1-29
3323	3323.757	0.756	MNTSHLLALLLPKSPQGENRSKPLGTPYNF(S)	1-30
4554	4554.221	0.221	MNTSHLLALLLPKSPQGENRSKPLGTPYNFSEHCQDSVDVM	1-41

Table S11: Screening of suitable detergents for enzymatic cleavage of the N-Y-KSI fusion peptides:

name	characteristics	concentration	100mMNaCl added
n-Decanoylsucrose	nonionic	9mM cannot solubilze the peptide	Improved a little
dihexanoylphosphatidylcholine	zwitterionic	4mM cannot solubilze the peptide	
n-Octyl-β-D-glucopyranoside	nonionic	35mM cannot solubilze the peptide	
n-Dodecyl-β-D maltopyranoside	nonionic	1-6x CMC cannot solubilze the peptide	
Fos-choline	zwitterionic	1-6x CMC cannot solubilze the peptide	Improved a little
n-Octyl-β-D-glucopyranoside	nonionic	1-6x CMC cannot solubilze the peptide	
n-Nonyl-β-D-maltoside	nonionic	1-6x CMC cannot solubilze the peptide	
n-Decyl-β-D-glucopyranoside	nonionic	1-6x CMC cannot solubilze the peptide	
Sarcosyl	anionic	0.4% can solubilze the peptide	completely soluble

Figure S12: Titration of N-Y2 with 0.5 (yellow), 1 (orange), 2 (red), 4 (purple), and 10 (blue) equivalents of pNPY.



Chapter 7

Expression of double transmembrane fragments of Y4 receptor, a human GPCR, for structural studies by NMR spectroscopy

7.0 Introduction

The Y4 receptor, a class A G-protein coupled receptor (GPCR) primarily targeted by the pancreatic polypeptide (PP), is involved in a large number of physiologically important functions. It plays a pivotal role in cardiac function, glucose metabolism in chronic pancreatitis patients and mediation of intestinal absorption of electrolytes and water[1].

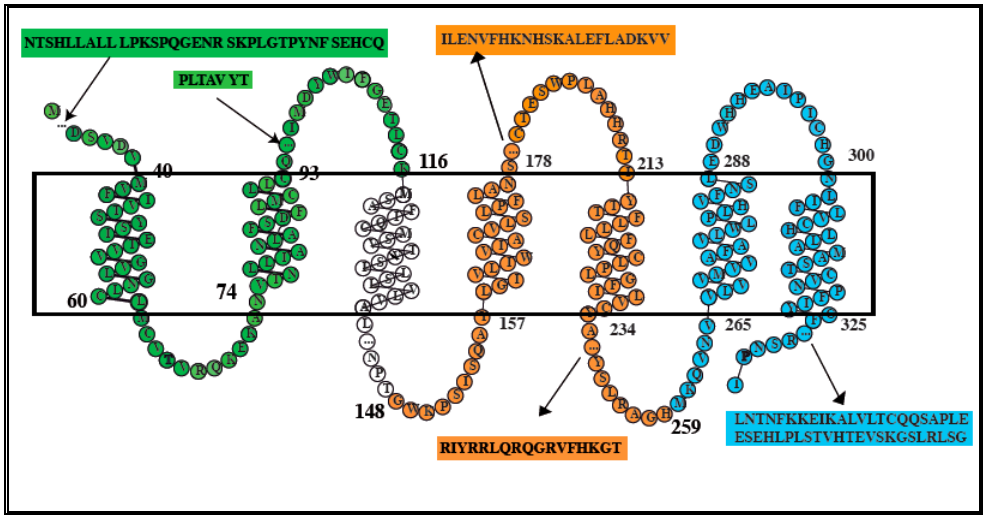


Figure 1:‘‘Snake’’-plot type presentation of the human Y4 receptor. The plot was modified from a download from the GPCR.org website. The parts of the receptor that has been expressed in this work are shaded in different colours. Y4 N-TM1-I1TM2-E1(116) (green coloured circles), Y4 I2(147)TM4-E2-TM5-I3(259) (orange coloured circles) and Y4 I3(260)TM6-E3-TM7-C (blue coloured circles). The numbering used follows that of the intact receptor. The omitted sequences for parts of the N terminus, C terminus and the loops are indicated separately in the box.

Our understanding of GPCR structures at the molecular level has been considerably improved over the last few years. However, it still remains a formidable challenge to obtain high-resolution structural data for intact GPCRs. Determination of membrane protein structure has proven a more difficult subject than the determination of soluble protein structure largely due to the notorious difficulty in achieving reasonable expression levels of active receptor and also due to the hydrophobic nature of membrane proteins and their resulting insolubility in aqueous medium. The recent development of new expression systems, mutant cell strains, and the availability of new lipids for membrane protein reconstitution and the effective methods for both selective and uniform labeling of proteins have opened the field of structural biology to many membrane proteins that were previously not readily accessible for examination. These advances, together with the parallel development of new pulse sequences, for both solution and solid-state NMR, provide a powerful means to tackle the problem of GPCR structure determination.

In this chapter we describe about the expression and purification strategy for the production of double transmembrane fragments I2(147)TM4-E2-TM5-I3(259) and I3(260)TM6-E3-TM7-C of the Y4 receptor in isotopically enriched form in *E. coli*. Both the double transmembrane constructs were successfully expressed as a TrpΔLE fusion protein, from which it was liberated by cyanogen bromide (CNBr) cleavage or by enzymatic cleavage in the presence of detergents using C3 protease and purified by RP-HPLC. Expression yields were optimized using different strains and growth temperatures as well as various induction parameters. The cleavage conditions were optimized. Extensive detergent screening was done to yield reasonable quality of spectra but still requires final optimization which is required for our analysis. Spectra with good quality could only be obtained when working under reducing conditions to eliminate fragment oligomerization through formation of intermolecular disulfide bonds. These experiments set the stage for the determination of the 3D structures of these large domains of GPCR in micelles using high resolution NMR.

There is no universal expression system suitable for functional production of all GPCRs, and the right combination between the target receptor and expression host must be found empirically. The *E.coli* expression system is of the first choice, because of its simplicity and cost effectiveness. Constructing expression vectors and trying out different expression strategies can be accomplished with relative ease, as compared to eukaryotic systems for which more sophisticated methods and a more expensive maintenance are required. Mutant proteins can also be rapidly generated and characterized. There are two strategies for the expression of GPCRs in *E.coli*: (a) expression of functional, membrane-inserted receptors[2-4] (b) expression of incorrectly folded, aggregated protein and refolding from inclusion bodies[5]. In both expression strategies, two types of expression methods were used. A direct expression, often used for the expression of large proteins offers the advantage that the chemical/enzymatic cleavage step is skipped a step that is often time-consuming and inefficient cleavage often results in low overall yields. Unfortunately this expression has been hampered by the relatively low yields of GPCRs owing to the toxic effects caused by these 7TM receptors when inserted into bacterial membranes. To circumvent this problem several strategies has been developed. Notably, fusing target proteins to a highly expressed bacterial protein has proven particularly effective for improving the expression level of membrane proteins in *E. coli[6]*. The type of fusion protein can be divided into two categories: soluble fusions and insoluble fusions.

Soluble fusions, as indicated by the name, can increase the solubility of the whole fusion protein and thus simplify the subsequent purification. In case of membrane proteins were highly hydrophobic transmembrane helices are present, the relatively large soluble fusion proteins like maltose binding protein (MBP) and glutathione S transferase (GST) are preferred provided that the fusion protein can be expressed in soluble form. After purification of the soluble fusion protein, removal of the soluble fusion is essential, and in most cases this is accomplished by enzymatic cleavage. The problem with the soluble fusion approach is that it is difficult to predict whether the fusion protein will be soluble or not. If the fusion protein is insoluble, tag removal then adds another layer of complexity, and low cleavage efficiency is often observed.

Insoluble fusion partners have been widely used to isolate double transmembrane domains. In such a system usually a very hydrophobic protein is used as the fusion partner like the ketosteroidisomerase (KSI) or TrpΔLE leader peptide fusions. Since the final fusion protein is more hydrophobic, it will most likely accumulate as inclusion bodies and increase the final yield. Even though the insoluble fusion is a robust method, a potential problem is related to the cleavage of the fusion. Since the fusion protein is not soluble in most of the enzyme-compatible solutions, cyanogen bromide cleavage in guanidine chloride/0.1M HCl in most cases is employed[7]. This requires that no methionine should be present in the target sequence. Alternatively proteases can be used to cleave the fusion partner. The most important point of concern in this strategy is that for many membrane proteins to be soluble at useful concentrations they require more acidic or more basic pH levels, high or low salt levels, or the presence of detergents or chaotropes. It is therefore essential that the protease of choice retains substantial activity under these conditions. The most robust of the enzymes cited appear to be the SUMO proteases, the profinity enzyme, TEV protease and the C3 protease. We have used the C3 protease, which retains activity in presence of detergents, to cleave the fusion partner. The affinity fusion tags like HIS, CBP, CYD (covalent yet dissociable NorpD peptide), Strep II, FLAG, HPC (heavy chain of protein C) peptide tags, and the GST and MBP are widely used as protein purification tags.

Fusion technology continues to expand with the introduction of new fusion partners, purification and detection tags, and cleavage reagents. Because every protein is unique, no single tag or cleavage method will answer every need. When considering which tag to use, key questions that should be asked are, whether our application tolerates retention of the tag or the presence of a few amino acids remaining at the cleavage site, requirement of expensive proteases, incomplete cleavage of the fusion protein and the requirement of additional steps of purification. Apart from the problems associated with the difficulties in expression and purification, NMR analysis is complicated by the large size of the GPCR /micellar complex and the poor chemical shift dispersion as well as the high frequency of occurrence of certain residues (such as Ala, Leu, Ile, Phe) in the

transmembrane region of the receptor. Therefore a critical need arises to look for other alternative approaches. According to Popot and Engelman the transmembrane domains can be thought of as independent folding units and can be studied separately[8]. It is therefore hypothesized that polypeptides or small proteins containing the amino acid sequences for turns or for transmembrane helices of membrane proteins built of helical bundles will exhibit the secondary structures in solution similar to that of the native protein. Use of peptides to determine secondary structure of membrane proteins is becoming widely accepted. A significant number of studies provided structural information on fragments corresponding to sequences of GPCR helices, loops and termini. E.g.: *Saccharomyces cerevisiae* pheromone receptor[9], cannabinoid receptor[10], and Neurokinin-1 receptor[11]. Significant similarities between the structures of fragments of rhodopsin and the corresponding regions of the x-ray crystal structure of native rhodopsin were observed[12].

Furthermore, since the interactions of the lipid bilayer with membrane proteins are crucial for its solubility and stability, it is desirable to determine protein structures in a good membrane mimetic environment. Nuclear magnetic resonance (NMR) spectroscopy is very well suited for this goal, since it can be applied to molecules in all physical states, including liquid crystalline bilayers and micelles formed by the lipids that associate with membrane proteins. Solubilization and stabilization of membrane proteins using detergents constitute a bottleneck for the structural biology of membrane proteins[13, 14]. Detergent micelles provide such environments surrounding the hydrophobic domains of membrane protein and keeping them soluble in an aqueous environment, and thus are widely used for solubilization and purification of membrane proteins[15]. Generally, the suitable detergent or detergent mixture should solubilize the target protein most effective, keep it stable, and prevent its self-aggregation. However, due to individual differences between membrane proteins, the choice of detergent or detergent mixture for a particular protein cannot be predicted. Therefore, a systematic approach is required to select the optimal detergents to achieve solubilization and stabilization of each target protein.

A number of additional factors contribute to the quality of the NMR spectra obtained from peptides and proteins in micelles. Optimal salt concentration, pH, temperature and lipid concentrations much higher than the critical micelle concentration are essential for preparing samples that yield NMR spectra with sufficiently narrow line widths and the appropriate number of resonances.

7.1 Materials and method

The pLC01 plasmid which contains TrpΔLE gene was obtained as a generous gift from Prof. Fred Naider (College of Staten Island, CUNY, NY 10314). All phospholipids were from Avanti Polar Lipids (www.avantilipids.com), octyl-glucopyranoside (OG) was from Fluka (www.sigmaaldrich.com), and deuterated sodium-dodecyl-sulfate (SDS) and dodecyl-phosphocholine (DPC) were from Cambridge Isotopes Laboratories (www.isotope.com). ^{15}N-NH_4Cl and D_2O were obtained from Cambridge Isotopes Laboratories (www.isotope.com). C3 Protease expression plasmid was obtained from the EMBL. Oligonucleotide primers were synthesized by Microsynth GmbH (Balgach, Switzerland).

7.1.1 Cloning and mutagenesis to generate double transmembrane constructs of human Y4 receptor into pLC01 vector

To generate a plasmid encoding the double transmembrane constructs TM4-TM5 and TM6-TM7C, with and without fusion protein the corresponding DNA sequences were subcloned from wt cDNA of the Y4 receptor (University of Missouri-Rolla (UMR) cDNA Resource Center) by PCR. For non-fusion double transmembrane constructs, the primer sequence contain *NdeI* and *BamHI* restriction sites and the PCR product was digested with *Nde1* and *BamHI*, the fragments were ligated with T4 DNA ligase into the pLC01 vector, which had also been digested with *Nde1*and *BamHI* to generate the double transmembrane constructs Y4 I2(147)TM4-E2-TM5-I3(259), Y4 I3(260)TM6-E3-TM7-C without TrpΔLE fusion protein.The target sequence pattern in pLCO1vector for the non –fusion construct is as follows

TM4-TM5 : Nde1site–Histag-Y4 I2(147)TGWKPSISQAY-TM4-E2-TM5-I3(259)-stopcodon-BamH1site

TM6-TM7-C(340): Nde1 site–Histag-Y4 I3(260)MKQVNV-TM6-E3-TM7-C(340)–stopcodon BamH1site

To generate the TrpΔLE fusion of the above mentioned double transmembrane constructs, in case of TM4-TM5 a Met residue was inserted at the N-terminus of the target sequence to remove the TrpΔLE fusion by CNBr cleavage. In case of TM6-TM7C construct a C3 protease cleavage site is introduced since it contains internal methionine residues in the target sequence. The primer sequence contain *HindIII* and *BamHI* restriction sites and the PCR product was digested with *HindIII* and *BamHI*, the fragments were ligated with T4 DNA ligase into the pLC01 vector that contains the TrpΔLE sequence, which had been digested with *HindIII* and *BamHI.* The generated plasmid encoded the TrpΔLE leader sequence fused to the Y4-I2(147)TM4-E2-TM5-I3(259) or Y4-I3(260)TM6-E3-TM7-C double transmembrane constructs.The sequence arrangement of both the constructs are represented as follows.

TrpΔLE-TM4-TM5: Nde1 site-Histag-TrpΔLE-(CNBr Cleavage site -Met) Y4 I2(147)TGWKPSISQAY-TM4-E2-TM5-I3(259)-stopcodon-BamH1site

TrpΔLE TM6-TM7-C: Nde1 site–Histag- TrpΔLE-(C3 protease cleavage site -LEVLFQ/GP)-Y4 I3(260)MKQVNV-TM6-E3-TM7-C(375)–stopcodon-BamH1site

We wanted to test using the N-terminus of Y4 receptor as a fusion partner. To produce this fusion plasmid the pLC01 vector containing the gene I2(147)TM4-E2-TM5-I3(259) without the TrpΔLE fusion was used as a template for site-directed mutagenesis. The DNA sequence corresponding to the N-terminus of the Y4 receptor was amplified from the wt cDNA by PCR. This insert was used as an megaprimer to amplify the plasmid which already contains the gene for TM4-TM5 double transmembrane region to generate an fusion Y4N-TM4-TM5 by site directed mutagenesis based on the Stratagene QuikChange mutagenesis protocol (Stratagene, La Jolla, CA). The reaction mixture was subjected to digestion (80 units, at 37°C) of *Dpn*I (NEB), which specifically cuts the methylated template DNA, and then transformed into DH5α competent cells (Invitrogen). The final plasmid pLC01generated, coded for Y4N-TM4-E2-TM5-I3(259). A slight difference exists in sequence between N-terminal fusion and TrpΔLE fusion double TM constructs. In the N-terminal fusion construct the I2 loop residues (10 amino acids) are not included and immediately after the N-terminal domain fusion, the 4th

transmembrane sequence starts.The sequence arrangement of N-terminal fusion of TM4-TM5 construct in pLC01 vector is as follows and for comparison TrpΔLE –TM4-TM5 sequence is also shown.

N-TM4-TM5: HindII site-Histag- Y4 N(1-41)- (CNBr Cleavage site -Met) (158)TM4-E2-TM5-I3(259)-stopcodon-BamH1site.

TrpΔLE-TM4-TM5: Nde1 site-Histag-TrpΔLE-(CNBr Cleavage site -Met) Y4 I2(147)TGWKPSISQAY-TM4-E2-TM5-I3(259)-stopcodon-BamH1site.

The correctness of all the constructs was verified by DNA sequencing (Synergene Biotech, Switzerland). For the sake of simplicity the fusion proteins will be mentioned as N-TM4-TM5, TrpΔLE-TM4-TM5 and TrpΔLE-TM6-TM7-C, and the cleaved double transmembrane fragments as Y4-TM4-TM5 and Y4-TM6-TM7-C here afterwards.

7.1.2 Protein expression and purification of the double transmembrane fragments

The plasmid encoding the target protein was transformed into BL21-AI cells for expression, which due to their very low background expression levels were previously shown to result in higher expression levels compared to other strains[7]. A freshly transformed colony was used to inoculate 10 ml LB containing 100 mg/ ml ampicillin. This preculture was grown over night at 37°C and then used to inoculate 1L LB (for the unlabeled sample) or M9 (with ^{15}NH4Cl and ^{13}C glucose as sole nitrogen and carbon sources) media containing 100 mg/ml ampicillin and cultured at 37° C until the OD600 reached 0.45–0.5. For induction of overexpression, the temperature was lowered to 20°C and 0.2% L-arabinose was added. Cells were harvested after 18-20 hrs and stored at -20°C until further use. The cell pellet was resuspended in binding buffer (50 mM Tris–HCl, and 500 mM NaCl, 5 mM β-mercaptoethanol, pH 8.0) and lysed by sonication. The resultant suspension was spun at 13,000 rpm for 30 min, the supernatant was discarded, and the solid pellet was resuspended at a ratio of 15 mL per liter of culture in binding buffer containing 6 M GdmCl and left to solubilize overnight at room temperature. After centrifugation (45 min at 19,000 rpm and 4°C) the supernatant was loaded into a 5 mL Hi-Trap chelating chromatography column (Amersham-Pharmacia Biotech), which had

been previously charged with 50 mM $NiSO_4$ and equilibrated with binding buffer containing 6 M GdmCl. Impurities bound to the resin were removed by passage of enough buffer (30 mM imidazole, 50 mM Tris–HCl, 500 mM NaCl, and 6 M GdmCl, 5 mM β-mercaptoethanol, pH 8.0) through the column until the A_{280}nm remained constant as monitored with the detector of an FPLC apparatus. The fusion protein was finally eluted from the affinity column with ~20 mL of elution buffer containing 300 mM imidazole, 50 mM Tris–HCl, 500 mM NaCl, 6 M GdmCl, β-mercaptoethanol, pH 8.0. Fractions containing fusion protein were dialyzed against pure distilled water (3-4 kDa) and the precipitate was recovered by centrifugation of the dialysate and lyophilized. The masses of the proteins were confirmed by MALDI–TOF MS. For Y4N-TM4-TM5 (mass expected-18405.0 Da; mass obtained-18401.2 Da. If we consider the 4 Cys in the oxidized form then the mass difference is 0.2 Da) and for TrpΔLE-TM6-TM7-C (mass expected -27076.4 Da; mass obtained -27069.9 Da. Again, if we consider 4 Cys in oxidized form then the mass difference is acceptable (2.5 Da).

The lyophilized protein was then subjected to CNBr cleavage in case of the TrpΔLE-TM4-TM5 construct and to a C3 enzymatic cleavage for the TrpΔLE-TM6-TM7-C construct (*vide infra*). The mixture was dialyzed (MWCO 3-8 kDa) against binding buffer containing 6M GdmCl for 4-5 hours and Ni-affinity chromatography under denaturing conditions was carried out to remove the His-tagged TrpΔLE peptide and also the C3 protease. The protein sample was reduced with 100 mM DTT, 250 mM mercaptoethanol and the target proteins were finally purified by C4-RP-HPLC (Vydac, USA) by using a water/acetonitrile/0.1% TFA gradient. The final yield of the target protein was found to be 1- 2mg/L.

7.1.3 CNBr cleavage

The lyophilized protein was dissolved in 6M GdmCl/0.1M HCl to a final concentration of 10 mg/ml. 1.5 times weight excess of solid CNBr was added to the solution, the reaction mixture was flushed with nitrogen for 3 min, then wrapped with foil and the reaction mixture incubated in the dark for 20-24h at room temperature. The mixture was dialyzed (MWCO 3-8 kDa) against pure distilled water for 4-5 hours. The

utilization of chaotropic agents in the presence of diluted acids is the preferred cyanogen bromide cleavage medium of fusion proteins in order to maximize cleavage efficiency of hydrophobic sequences and to prevent deleterious degradation and structural modifications of the target peptides. The efficiency of the cleavage was around 75-80%. All these procedures were conducted under a fume hood.

7.1.4 Production of N-His_6-3C protease

The plasmid encoding the target protein N-His6-3C protease in a pET-24d (+) vector was transformed into BL21-DE3 cells for expression. A freshly transformed colony was used to inoculate 10 ml of LB containing required amount of 50 mg/ml kanamycin. This preculture was grown over night at 37°C and then used to inoculate 1L LB containing 1ml of 50mg/ml kanamycin culture until the OD600 is approx. 0.6. Then the culture was cooled on ice to 20°C. Over expression was induced by the addition of 0.2 mM IPTG after which the culture was grown overnight at 20°C. Cells were harvested and the cell pellet from 1 L culture was resuspended in lysis buffer (50 mM Tris–HCl, and 200 mM NaCl, pH 8.0) and sonicated to lyse the cells. The supernatant of the cell lysate containing the recombinant C3 protease protein tagged with 6x Histidine was applied to pre-equilibrated Ni-NTA resin (Sigma). The protein-bound to Ni-NTA resin was washed with chelating buffer (50 mM Tris–HCl, and 300 mM NaCl, 1 mM β-mercaptoethanol, 10%glycerol, pH 8.0) containing 30 mM imidazole. The bound proteins were eluted with chelating buffer containing 300 mM imidazole. The eluted protein was then dialysed against storage buffer (50 mM Tris–HCl, 150 mM NaCl, 1mM DTT and 20% glycerol, pH 8.0). The yield was found to be 300mg/L of LB medium.

7.1.5 Screening of cleavage conditions for TrpΔLE-TM6-TM7-C by the C3 protease

A mixture of the HRV-3C Protease and the target protein of 1:5(unit/μg) were added to 20mM Tris buffer containing 150mM NaCl and various concentrations of different detergents. (A concentration range of 0.5% and 1% was used for Triton-X-100, Tween-20, CHAPS, N-lauroyl sarcosine, *N*-lauroyldimethyl amine oxide (LDAO), and

0.01%, 0.05% and 1% for DPC and LPPG were used). The reaction mixture was incubated at 4°C for 16 hrs and the extent of cleavage of the protein samples determine by SDS-PAGE analysis. An undigested sample of the target protein was run as a control.

7.1.6 NMR sample preparation

Uniformly labeled [^{15}N]-N-TM4-TM5,(fusion protein) [^{15}N]-TM4-TM5 protein was dissolved in 200 µl 90% H_2O/D_2O containing 5% of respective detergents or phospholipids and 40mM DTT by thorough sonication for 30 min. 50µl of 0.2 mM phosphate buffer (pH 6.0) was added to give a final concentration of 40mM phosphate. The peptide/detergent mixtures were then transferred to a Shigemi NMR tube. Nitrogen-proton heteronuclear single quantum correlation ([^{15}N,^{1}H]-HSQC) NMR spectra were recorded on an AV-600 Bruker NMR spectrometer at 320 K. Chemical shifts of protons were calibrated according to the water line at 4.53ppm, from which the nitrogen chemical shifts were referenced indirectly using the conversions factors published on the BMRB database. The spectra were processed using the Bruker Topspin 2.0 software.

7.2 Results and discussion

7.2.1 Design and generation of double transmembrane constructs of human Y4 receptor into pLC01 vector

Previously Dr.Chao Zou from our group has studied the double transmembrane fragment of the Y4 receptor (N-TM1-TM2) comprising the N-terminal domain, the first two transmembrane (TM) helices and the first extracellular loop followed by a $(His)_6$ tag. Almost complete assignment of the backbone, including all TM residue resonances was obtained. We further wanted to validate the approach of using fragments of GPCRs as surrogates to probe receptor structure. We decided to generate double transmembrane fragments corresponding to the region TM4-TM5 and TM6-TM7C to determine their structure in membrane-mimicking environments (Figure 1). Due to their smaller size or due to their toxicity to the *E.coli* cells, direct expression was not possible. Therefore these transmembrane fragments were designed as TrpΔLE fusion proteins and liberated from their fusion partner using cyanogen bromide cleavage or by C3 protease cleavage. It is

generally assumed that a disulfide bond between extracellular loop (E- I and E- II) is formed in family-1 GPCRs. Since this would further complicate the synthesis and the analysis we abstained from making the double transmembrane fragments corresponding to the TM3-TM4 domains. Since the ligand binding site of Y4 receptor involves the N-terminal part and the extracellular loops, it will be useful to study the interaction of ligands with TM4-TM5. Furthermore, TM6 of GPCRs have been shown to be extremely important in signal transduction[16, 17].

7.2.2 The choice of the expression vector systems

Critical factors governing efficient expression of any protein in *E.coli* include an adequate promoter in terms of strength, leakiness and inducibility, a low plasmid copy number if the target protein shows potential toxicity for the host bacteria, as well as a resistance marker that minimizes plasmid loss under strong counter selective pressure. Attention should also be given to good positioning and accessibility of translational initiators and to the presence of transcription terminators that insure mRNA stability and prevent promoter read-through that would compete with transcription.

The pLC01 plasmid that is derived from pET vectors is one of the most powerful systems developed for the cloning and expression of recombinant proteins in *E. coli*. Target genes cloned in pET plasmids are under the control of the strong bacteriophage T7 promoter, which contains the T7 RNA polymerase-binding site and the expression is facilitated by transferring the plasmid into an expression host containing a chromosomal copy of the T7 RNA polymerase gene under lacUV5 promoter (BL21DE3,C41DE3) or pBAD promoter control(BL21-AI strain). It is also possible to attenuate the expression level simply by lowering the concentration of inducer or by lowering the post induction temperature.

The fusion contains the N-terminus leader peptide TrpΔLE derived from the tryptophan operon**.** The advantage of the trp leader fusion sequence is that it helps to promote high levels of expression, which cause the fused polypeptide to form insoluble aggregates (inclusion bodies) in the cellular host, from which it can be solubilized and

purified. The Trp ΔLE sequence starts after the *Nde* I site with the first residue of the trp L gene and continues through codon of Ala-105. This codon is followed by a *Hind* III site, a met codon (TM4TM5) or a /LEVLFQ/GP C3 cleavage site (TM6TM7C) and the gene for corresponding for double transmembrane fragments of the Y4-receptor. The gene of interest was cloned between and *Hind* III and *BamH I.* The 6 methionines in the TrpΔLE sequence were replaced with leucines to simplify the cyanogen-bromide cleavage reaction and the 2 cysteines were replaced with alanines[18]. For Y4-TM4-TM5, the N-terminal domain of Y4-receptor was also used as a fusion protein. The advantage of using the N-terminal domain fusion is that owing to its small size (41 amino acids) direct characterization might be possible without the need to remove the fusion protein.

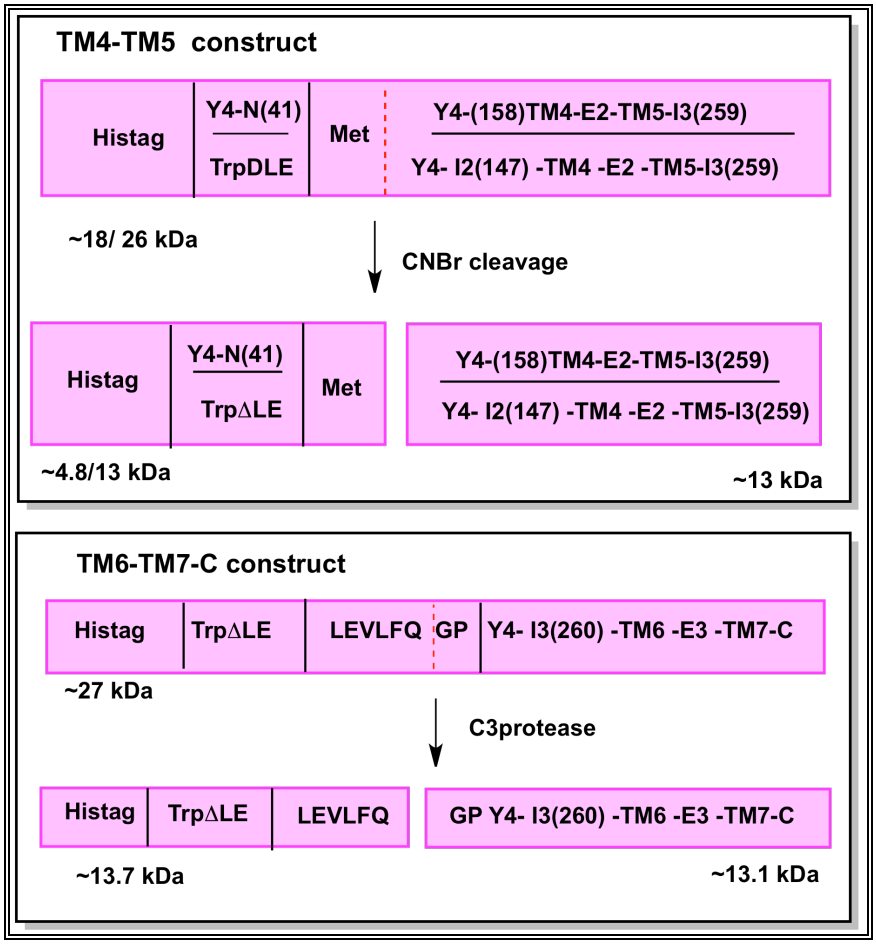


Figure 2: Scheme showing the two strategies used to produce double transmembrane fragments corresponding to the transmembrane regions of the Y4 receptor

In all the constructs a His–tag is present in the N-terminus of the fusion partner to facilitate purification by Ni-NTA affinity purification. In case of Y4 TrpΔLE-TM6-TM7-C construct following C3 protease cleavage two extra amino acids Gly and Pro are present along with the target sequence (Figure 2). In case of Y4TrpΔLE -TM4-TM5 and Y4N-TM4-TM5 constructs no extra residues of the vector origin was left over in the target gene after the removal of the fusion protein.

7.2.3 Selections of *E.coli* host strains

Although expression in *E.coli* is often simple for soluble proteins, major obstacles are encountered in the expression of many heterologous proteins especially for the integral membrane proteins. Unfortunately, many of these proteins severely interfere with the physiology of *E. coli*. As a result, expression yields are dramatically diminished, and sometimes no expression is observed at all. Therefore, modified expression vectors, modified *E. coli* strains, and appropriate cultivation protocols are needed. Various host vector combination has to be tested to obtain optimum yields.

A variety of *E. coli* strains designed to address many of these problems are commercially available. BL21(DE3) is the most widely used expression host. In *E. coli* strain BL21 (DE3), the gene encoding T7 RNAP is under the control of the poorly repressed *lacUV5* promoter. This leads to leaky basal expression in these strains which may be problematic when the target protein is toxic to the bacteria. Silent mutations that accumulate over several generations can be selected for their ability to "immunize" *E. coli* against toxic genes. Miroux and Walker described the selection of two derivative strains called C41(DE3) and C43(DE3) that tolerated the expression of some toxic genes[19]. In strain C41(DE3), the basal expression levels were five times lower than those of a standard BL21(DE3). In addition, after induction, the target gene product was found to accumulate more slowly. T7 RNA Polymerase (T7 RNAP) is not only a highly selective enzyme but also a highly efficient enzyme. Indeed, small amounts of T7 RNAP are sufficient to direct high transcription levels from T7 promoters located on multicopy plasmids[20].

Using the P_{BAD} promoter of the araBAD arabinose operon and the gene coding the positive and negative regulator of this promoter, *araC*, an induction-to-repression ratio of 1200:1 was obtained[21]. This principle is used in vectors containing P_{BAD} promoter to suppress the basal expression sufficiently. The benefits of the tightly regulated P_{BAD} promoter and the efficient T7 expression system have been now combined in one *E. coli* strain, in which a chromosomal copy of the T7 RNAP was put under the control of the P_{BAD} promoter. Because the T7 RNAP gene is inserted into the araB locus of the araBAD operon, expression of T7 RNA polymerase can be tightly regulated by the sugars L-arabinose and glucose. This arabinose-induced host (BL21-AI, Invitrogen) has 4-fold lower basal expression levels than BL21(DE3)pLyS and is especially useful to express genes that may be toxic to other BL21 strains where basal expression of T7 RNA polymerase is leakier.

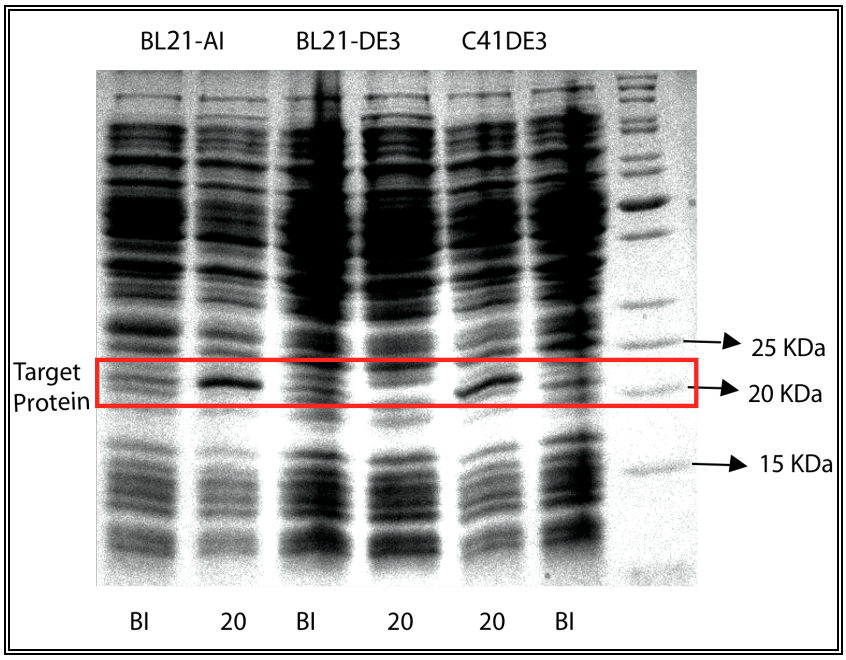


Figure 3: Selection of strain BL21-AI, BL21DE3, C41DE3 for expression of TrpΔLE-Y4 TM4-TM5. BI denotes ‘‘before induction’’ and 20 denotes ‘‘post induction growth temperature at 20°C’’

To evaluate the variation of our target protein expression level in different bacterial hosts three different strains BL21(DE3), C41(DE3) and BL21-AI were screened. The expression for the TrpΔLE-Y4 TM4-TM5 construct was tested for each strain at 20°C and the results were shown as an example in Figure 3. Considering the perfect control of leakage expression, and based on our previous successful experience for the construct Y4 N-TM1-TM2, BL21-AI was chosen as the more suitable host for large-scale expression; nevertheless the difference in yield in comparison to strains C41 (DE3) is not very significant.

7.2.4 Optimization of inducible expression and purification

Our goal was to utilize the high-level expression power of *E. coli* to produce the double-transmembrane constructs as inclusion bodies. Inclusion bodies can be an advantage for purification because 1) they are easily isolated by centrifugation to yield highly concentrated and relatively pure protein, and 2) inclusion body formation protects the protein from proteolytic attack. In addition, toxic proteins may not inhibit cell growth when present in inclusion bodies. However, these inactive, insoluble aggregates must then be solubilized and refolded to obtain fully folded receptor protein.

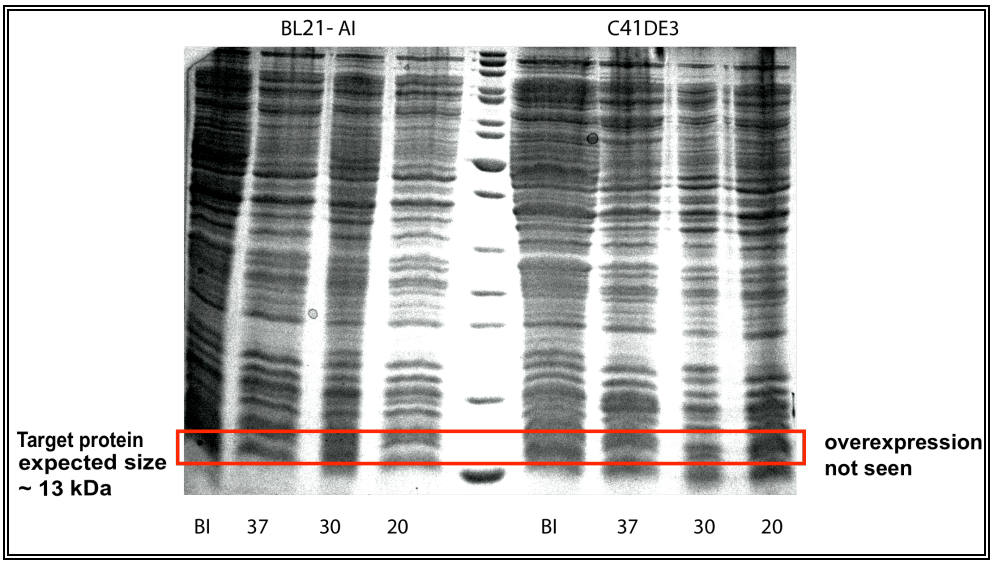


Figure 4: Screening of *E.coli* strains BL21-AI, C41(DE3) and temperatures for expression of Y4-TM4-TM5 (non fusion construct). BI denotes ‘‘before induction’’, 37, 30, 20 denotes ‘‘post induction growth temperature at 37°C, 30°C, 20°C respectively. There was no over-expression under any of the conditions (expected molecular weight is around 13 kDa).

Our previous attempt to express N-TM1-TM2 of the Y4 receptor by direct expression leads to reasonable amounts of pure protein required for its structural characterization[22]. Based on the success of this strategy and also to avoid the chemical/enzymatic cleavage and minimize the steps in purification, first we tried direct expression of the double transmembrane segments, (TM4-TM5; TM6-TM7) of the Y4 receptor without any fusion partner. We tested the three different host strains BL21(DE3), C41(DE3) and BL21-AI at different three temperatures 37°, 30° and 20°C with variable inducer concentrations. In none of these conditions we could see any over-expression. The expression screening with two different *E.coli* strains at different temperature for Y4- TM4-TM5 (non-fusion) is presented in Figure 4.

We therefore sought to utilize the fusion strategy to achieve better expression. We wanted to test using the N-terminus of Y4 receptor as a fusion partner as successfully done for the NY4-TM1-TM2 expression as a small fusion tag to favour the expression of double-transmembrane constructs. This fusion strategy takes advantage of the small size of the N-terminal domain, which may allow direct NMR analysis without the need of removing the fusion protein. In addition, the N-terminal domain was already assigned by our group[23] allowing to distinguish its resonances from those of the target protein. So to test this strategy an N-terminal domain fusion of TM4-TM5 was generated (Y4N-TM4-E2-TM5).

Since our collaborating group of Fred Naider successfully expressed double transmembrane constructs from the Ste2p receptor as TrpΔLE fusion proteins[7], we also cloned 2-TM constructs of TM4-TM5 and TM6-TM7 as TrpΔLE fusions along with the N-terminal domain fusion constructs. A reasonable expression level of around 3-4 mg/L of the fusion proteins was obtained for both the TM4-TM5 constructs (N-terminal domain fusion as well as TrpΔLE fusion) while a slightly higher yield of around 4-5mg/L for the TrpΔLE -TM6-TM7-C construct was achieved (Figure 5).

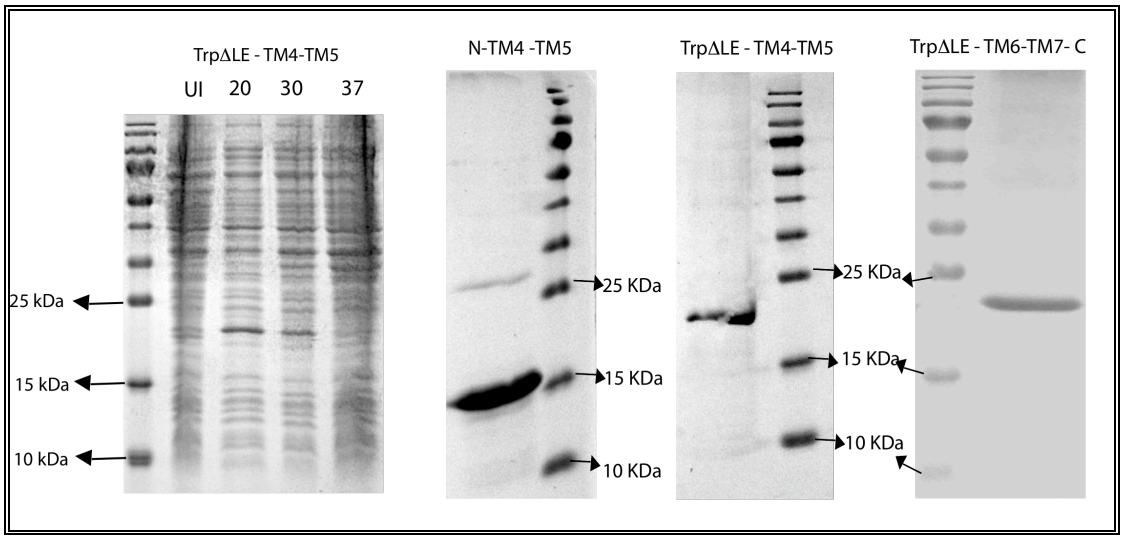


Figure 5: Screening of temperature conditions for expression of TrpΔLE Y4-TM4-TM5. UI denotes ‘‘uninduced’’, 37, 30, 20 denotes ‘‘post induction growth temperature at 37°C, 30°C, 20°C respectively (left). The target protein N-TM4-TM5, TrpΔLE TM4-TM5 and TrpΔLE TM6-TM7-C after Ni-NTA affinity purification (right).

As expression of GPCR fragments in *E.coli* is rarely trivial, modulating temperature, concentration of inducer or additives to the medium can significantly alter protein expression levels. It is well known that low temperatures will result in more protein expression, while higher temperature is detrimental because of the additional stress on the expression host. It is also known that proteolysis increases at higher temperatures. To configure the optimal temperature for over-expression, three different post induction growth temperatures (37°, 30° and 20°C) were tried (Figure 5). From the results of expression analysis it was found that post-induction growth at 20°C promoted the highest level of expression of the target protein. *E.coli* strains controlled by inducible promoters are also subjected to varying protein expression levels by altering concentrations of the inducing agent. As in our case we checked different amounts of arabinose with the BL21 AI strain, and different concentrations of IPTG for C41(DE3) and BL21(DE3). Finally, the use of M9 minimal medium supplemented with glucose (0.2–1%) instead of an LB rich medium can also improve expression of toxic genes, because glucose-mediated catabolite repression reduces the basal expression levels of T7 RNAP in bacterial host strains. From small-scale expression screening of various growth parameters it was found that usage of BL21-AI host strain with 0.2% arabinose inducer

concentration and induction growth temperature at 20°C present the most suitable expression conditions for large scale target protein expression.

The presence of contaminating lipids and low molecular weight proteins can introduce unwanted heterogeneity into the sample, broadening line widths or even introducing spurious resonances. Lipid–protein interactions may allow host cell lipids to remain bound to the protein throughout the Ni^{2+} affinity purification. In these cases further purification via RP-HPLC may be necessary to effectively remove the contaminants and improve sample homogeneity. We could previously demonstrate that removal of such lipids resulted in dramatically different solubility of the protein accompanied with much improved NMR spectra in case of the NY4-TM1-TM2 construct[22]. Low levels of high molecular weight protein contaminants, while undesirable, may be tolerated if they do not interact with the protein of interest since the low concentration and long rotational correlation times of these proteins typically prevent them from being observed by NMR.

7.2.5 Optmization of expression in deuterated growth medium

For structural studies of the membrane proteins, the concomitant increase in molecular weight as a consequence of micelle incorporation results in a dramatic decrease in spectral quality. In addition, slow conformational exchange processes lead to additional line- broadening. This has led to the frequently encountered experience that signals from the TM regions of membrane proteins remain invisible[24] . Thus deuteration is essential to yield spectra of sufficient quality. An small scale test expression was done to check for the overexpression of TrpΔLE-TM4-TM5 fusion protein in deuterated medium. To allow expression in deuterated water, plasmid containing the target gene TrpΔLE-TM4-TM5 transformed in BL21-AI cells were plated on a D_2O LB agar plate and incubated at 37°C for 18-22 hrs and one colony was used to inoculate a LB preculture in 100% D_2O containing 100 mg/ml ampicillin. The preculture was grown at 37°C overnight and was then used to inoculate 10 ml of 50% D_2O M9 containing 50 mg/ml ampicillin and incubated at 37°C (230 rpm) for 18-22 hrs. 2%of the inoculum was inoculated into 10 ml of 100% D_2O M9 containing 50 mg/ml ampicillin.

After incubation at 37°C, overexpression was induced when the OD_{600} had reached 0.45 by adding 0.2% L-arabinose at 20°C, and cells were harvested after 18 h. By increasing the level of deuteration sequentially we were able to see the overexpression (Figure 6).

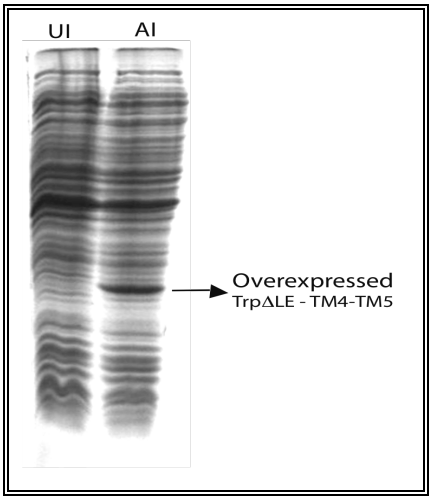


Figure 6: Small scale expression of TrpΔLE Y4-TM4-TM5 in fully deuterated M9 medium. UI denotes ‘‘uninduced’’ and AI denotes ‘‘after induction ’’ sample.

7.2.6 Optimization of the cleavage of the fusion protein

Once the protein is successfully expressed in a fusion system, it needs to be cleaved from the fusion protein. By incorporating a single methionine residue into the construct, the protein can be deliberated from the fusion protein using cyanogen bromide cleavage (Figure 7). An advantage to this is that the reaction is done under very acidic conditions, which is ideal for solubilizing these very hydrophobic transmembrane domains. But it requires that there are no internal methionines in the target sequence or that those can be removed without loss of structure/function by site-directed mutagenesis. In case of TM4TM5 there is no internal methionine in the sequence and we used this method to remove the TrpΔLE and N-terminal domain fusion.

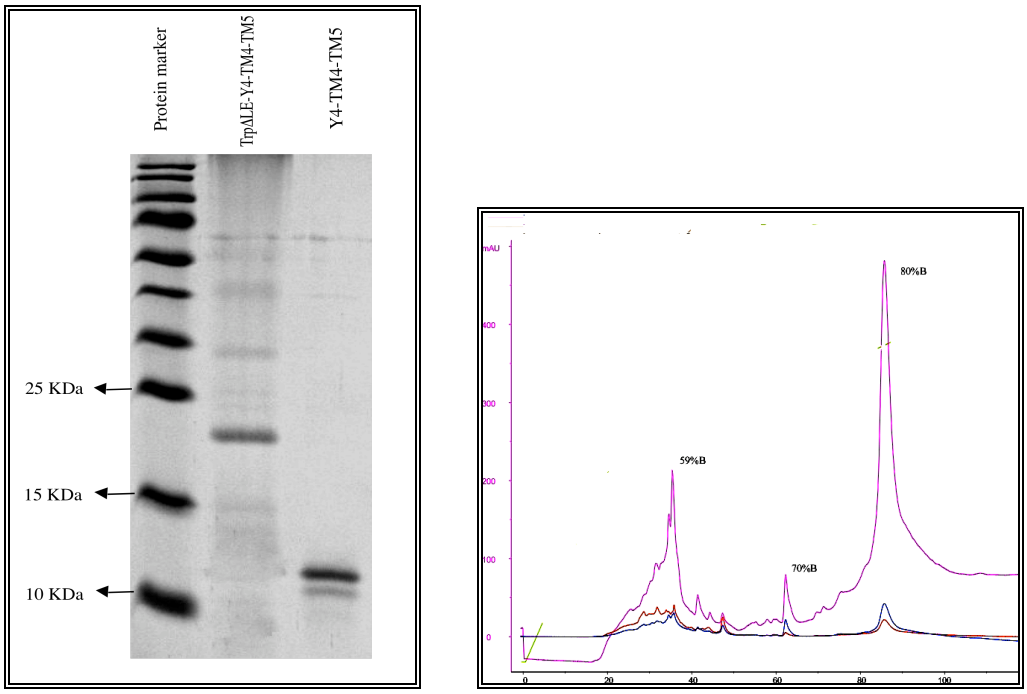


Figure 7: Cyanogen bromide cleavage of TrpΔLE-Y4 TM4-TM5 checked by SDS –PAGE electrophoresis (left). CNBr cleavage product subjected to C4 RP-HPLC purification, the peak at 80% acetonitrile corresponds to the cleaved target protein Y4-TM4-TM5 (right).

In TM6-TM7-C construct the C3 Protease cleavage site is introduced to facilitate the removal of TrpΔLE leader from the target protein. HRV 3C Protease is a recombinant form of the 3C protease from human rhinovirus type 14. A number of factors, including its small size (22 kDa), its robust activity at 4°C, high specific activity, activity in some detergent containing buffers and a His•Tag for its removal make HRV 3C protease an ideal choice for rapid cleavage of fusion proteins[25-27]. It recognizes the cleavage site: LEVLFQ/GP[26]. C3 protease was produced recombinantly in the BL21(DE3) strain and the yield was found to be 300mg/l of culture (Figure 8).

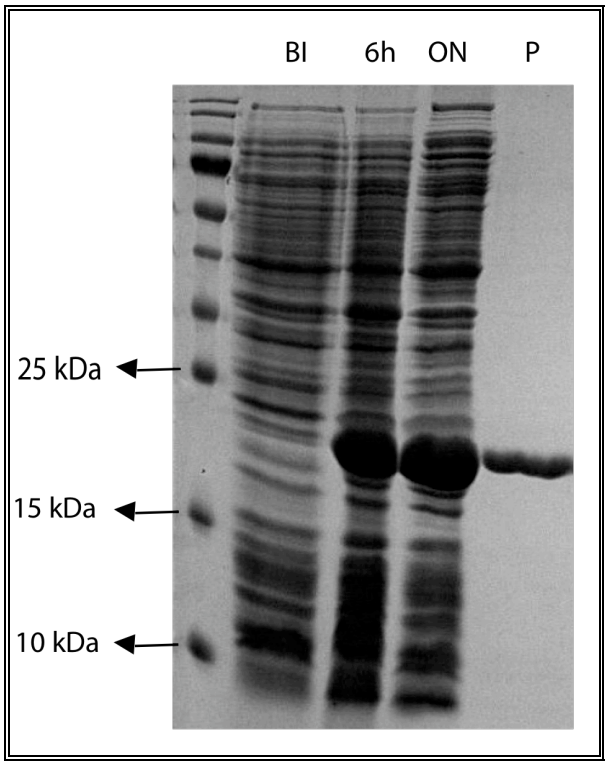


Figure 8: Expression of C3 precession protease in the BL21 (DE3) strain. BI denotes "before induction", 6h denotes 6hrs after induction, ON overnight induction and P denotes pure C3 protease protein after Ni-NTA affinity purification**.**

Since the cleavage site is accessible to different extent in different proteins we needed to test several enzyme-to-target protein ratios, different detergents and incubation times to optimize the efficiency of cleavage. Especially in case of hydrophobic transmembrane constructs we had to find a suitable detergent buffer which solubilizes the protein as well as retains the C3 protease activity. A preliminary screening with several detergents like Triton-X-100, Tween-20, CHAPS, N-lauroyl sarcosine, *N*-lauroyldimethyl amine oxide (LDAO), DPC and with phosopholipids like LPPG was done. We screened at 0.5% and 1% concentration for detergents and for DPC and LPPG 0.05%, 0.1% and 1% were used. The results of the detergent screening for optimizing 3C protease cleavage are summarized in the Table 1.

Detergent	Type	CMC mM	Removal by Dialysis	Solubility of the protein	Cleavage of the fusion protein
Triton-X-100	Nonionic	0.24	No	Very good	No
Tween -20	Nonionic	0.06	No	Moderate	Yes
CHAPS	Zwitterionic	8-10	Yes	Moderate	Yes
LDAO	Zwitterionic	2	Yes	Moderate	Yes
N-lauryl sarcosine	Anionic	*ca* 15	Yes	Very good	No
DPC	Zwitterionic	1.5	No	Moderate	Yes
LPPG	Zwitterionic	0.02–0.6	No	Very good	Yes

Table 1: Summary of results of the C3 protease cleavage of TrpΔLE Y4 TM6-TM7-C in different detergent buffers (1% detergent is used for all reactions except for DPC and LPPG for which 0.05% is used)

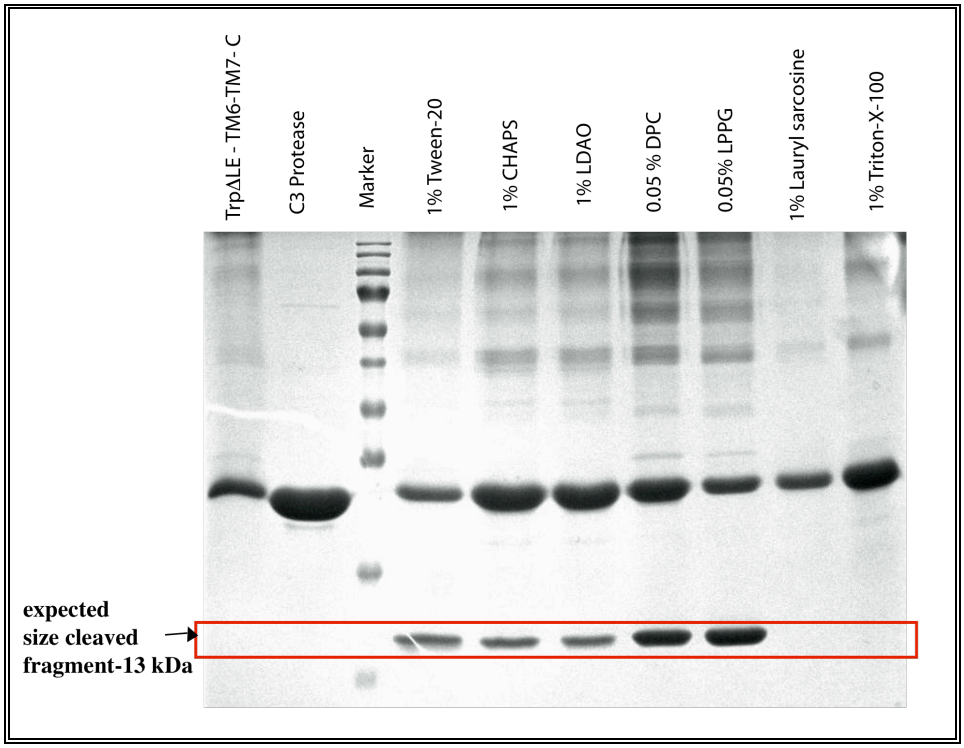


Figure 9: Summary of results of the C3 protease cleavage of TrpΔLE Y4 TM6-TM7-C in different detergent buffers (1% detergent is used for all buffer except for DPC and LPPG for which 0.05% is used) checked by SDS –PAGE gel electrophoresis

Considering the three important factors like solubility, cleavage efficiency and dialysability, CHAPS or LDAO were found to be most suitable (Figure 9). An enzyme-

to-target protein ratio (U/μg) of 1:5 was used for our cleavage reaction set up. Furthermore to optimize the cleavage efficiency and solubility of the target protein and to minimize the formation of higher molecular weight aggregates, varying concentrations of some buffer ingredients like NaCl and some buffer additives have to be tested.

7.2.7 Detergent screening

An ideal membrane mimetic for solution NMR will solubilize the integeral membrane protein (IMP) and allow for the acquisition of well-resolved spectra. Furthermore the protein should be stable enough to permit a series of multidimensional NMR experiments to be acquired on a single sample. Detergent micelles are the smallest membrane mimics that provide all of the interactions that integral membrane proteins require, prearranged in the appropriate order. It is not surprising that different detergents have such dramatic effects upon the spectra, as a number of complex interactions are involved. Thus selection of an appropriate detergent for NMR is a significant challenge. Furthermore the shape and dynamics of the detergent micelles can vary with pH and temperature thereby altering the environment surrounding the IMP. Spectra with broad line-widths and missing resonances are indicative of a micellar environment for which the protein is aggregated or has slow global motions resulting in poor spectral properties throughout the sample. Alternatively, the presence of more resonances than expected may indicate the presence of multiple conformations of the IMP or variations among the detergent micelles within the sample.

Since it is not currently possible to determine the best detergent *a priori*, a number of different detergent types including anionic (SDS and sarcosyl), long chain (LMPG and LPPG), medium chain (DPC), and short chain (DHPC) zwitterionic detergents should be screened to optimize NMR spectra. Detergent screens via analysis of a series of ([^{15}N,^{1}H]-HSQC) spectra are vital for selecting the appropriate detergent for each protein. The variation of resonance intensities within each spectrum can be a function of the detergent. These variations are likely the result of differing dynamics in spatially disparate regions of the protein or a result of sample heterogeneity.

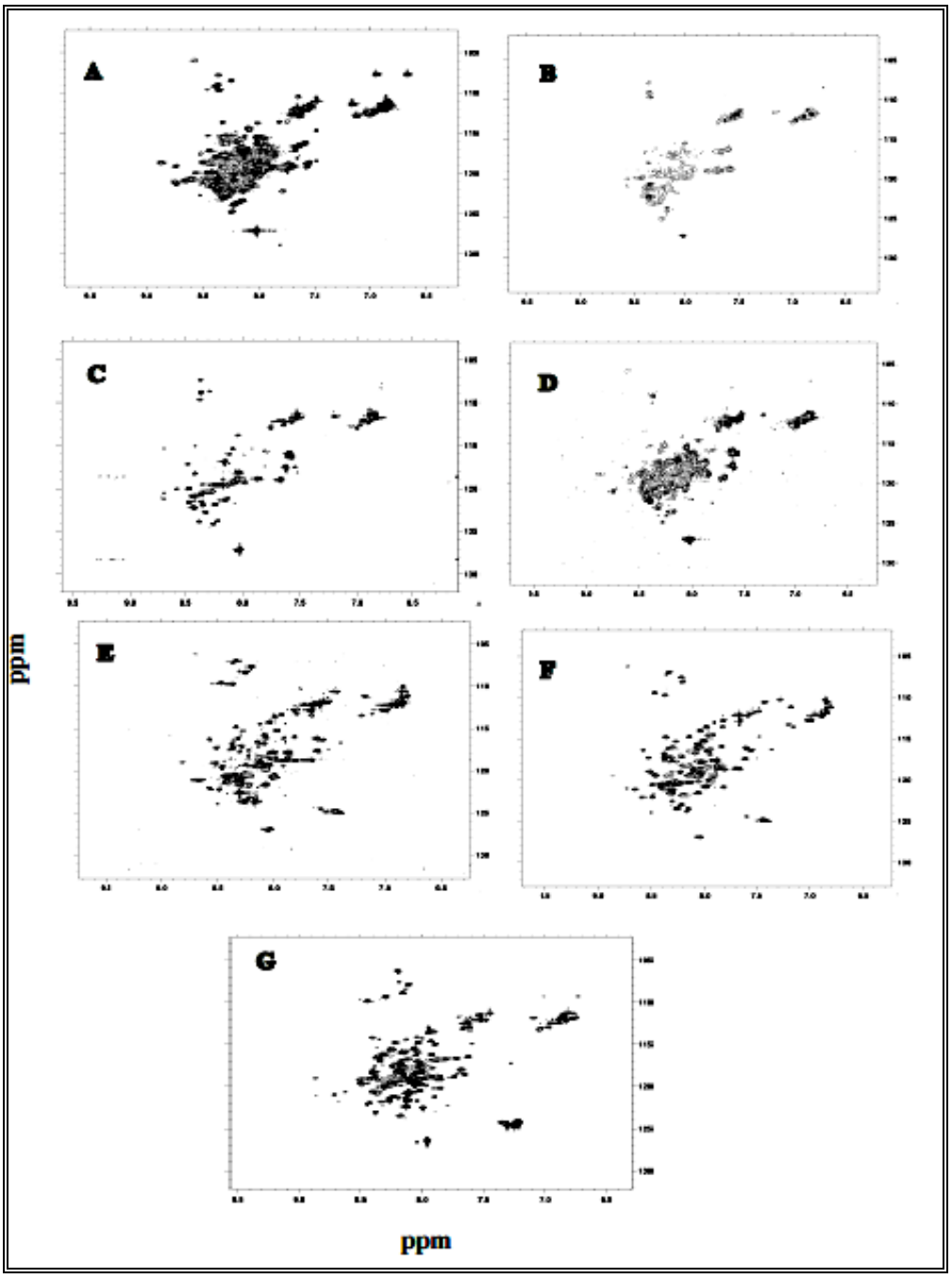


Figure 10: Plots of the two dimensional [^{15}N,^{1}H]-HSQC spectra of Y4-N-TM4-TM5. Spectra were recorded using 0.2 mM samples of the protein in 40mM phosphate buffer containing 40mM DTT at pH 6.0 in 5% LDAO(A); 5% OGP(B); 300mMDPC (C); 5% DHPC (D); 5%LMPG (E) ; 5% LPPG (F) and 170mM SDS (G). All data were recorded at 47°C at 600 MHz proton frequency. All the spectra were recorded with protein samples of highest purity and homogeneity after RP-HLPC purification

For the N-TM4-TM5 construct, we measured a series of ([^{15}N,^{1}H]-HSQC) spectra in the following detergents: DPC (n-dodecylphosphocholine), DHPC (1,2-dihexanoyl-*sn*-glycero-3-phosphocholine), LPPG (1-palmitoyl-2-hydroxy-*sn*-glycero-3-[phosphor-RAC-(1-glycerol]), LMPG (1-myristoyl-2-hydroxy-sn-glycero-3-[phospho-RAC-(1-glycerol)]), OGP (β-D octylglucopyranoside), SDS (sodium-dodecyl-sulfate) and LDAO. Though there is a drastic variation of the spectra in different detergents none of them resulted in spectra of sufficient quality to enable recording triple resonance experiments for the backbone assignment (Figure 10). The spectra measured in lysophospholipids (LPPG and LMPG) and SDS appears to be better dispersed compared to spectra measured in other detergents.

We also tried some detergent mixtures (e.g. LPPG and SDS; LPPG and DPC) but unfortunately no improvement in the spectral quality was observed. Then we decided to remove the N-terminal fusion domain by CNBr cleavage. After the removal of N-terminal domain fusion, the spectra measured with cleaved TM4-TM5 fragment shows peaks with broad line widths indicative of highly aggregated state of the protein. But the TM4-TM5 fragment cleaved from the TrpΔLE peptide using CNBr to our surprise resulted in good quality spectra in LPPG and SDS micelles (Figure 11). This could be due to the slight difference in their sequence. In the N-terminal fusion construct, the I2 loop residues (10 amino acids) are not included and immediately after the N-terminal domain fusion, the TM4-TM5 sequence starts. This sequence begins with Tyr which might lead to the pi-stacking interactions among the double TM fragments. Along with this, the highly hydrophobic nature of double TM and the formation of disulphide bonds among the transmembrane fragment might lead to a highly aggregated state of the protein. But the TM4-TM5 fragment generated from the TrpΔLE fusion has residues from the I2 loop regions before the actual TM region begins. This might be helpful in minimizing the formation of aggregates. This might be one of the reason for the better resolved spectra for the double TM fragment I2(147)TM4-E2-TM5-I3(259)(cleaved from TrpΔLE fusion) than the (158)TM4-E2-TM5-I3(259) (cleaved from N-terminal domain) which does not contain the I2 loop but further sample optimization is needed to obtain well dispersed spectra.

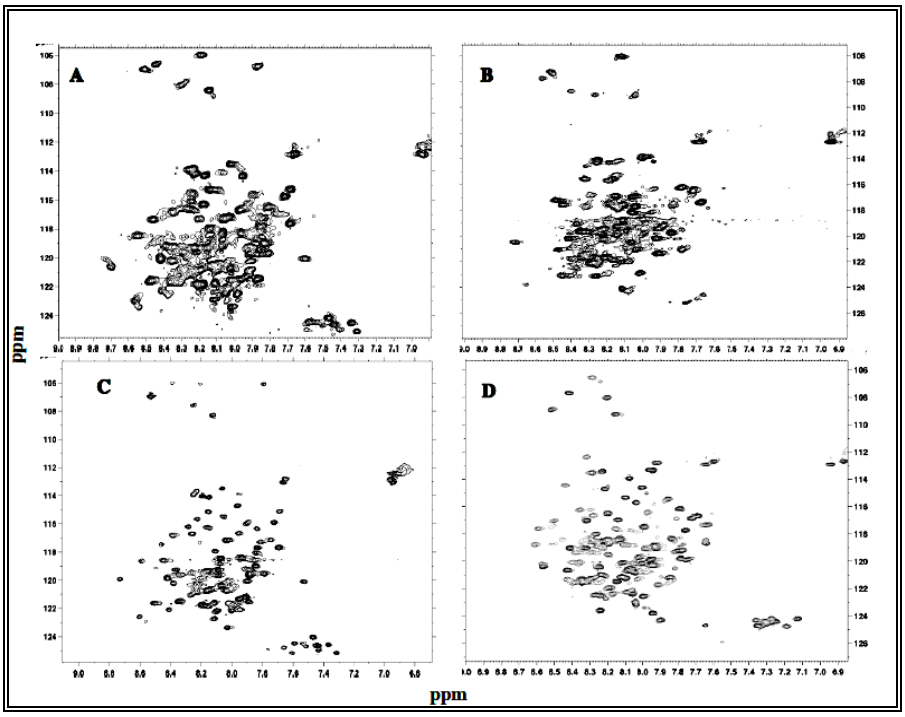


Figure 11: Plots of the two dimensional [^{15}N,^{1}H]-HSQC spectra of Y4-TM4-TM5 purified after CNBr cleavage from TrpΔLE-Y4-TM4-TM5. Spectra were recorded using 0.2 mM samples of the protein in 40mM phosphate buffer containing 40mM DTT at pH 6.0 in (A) 300mM DPC; (B) 5% LMPG (C); 5% LPPG and (D) 170mM SDS. All data were recorded at 47°C at 600 MHz proton frequency. All the spectra were recorded with protein samples of highest purity and homogeneity after C4 RP-HLPC purification.

It is common that a detergent performs very well under one set of conditions, yet poorly at a different pH or temperature. In case of the Y4-N-TM4-TM5 construct we have measured an series of [^{15}N,^{1}H]-HSQC spectra at different pH (pH5.0, 5.5, 6.0, 6.5, 7.0 and 7.5) in LPPG micelles (Figure 12). At slightly lower pH range (5.0 -6.0) spectra appears to be better.

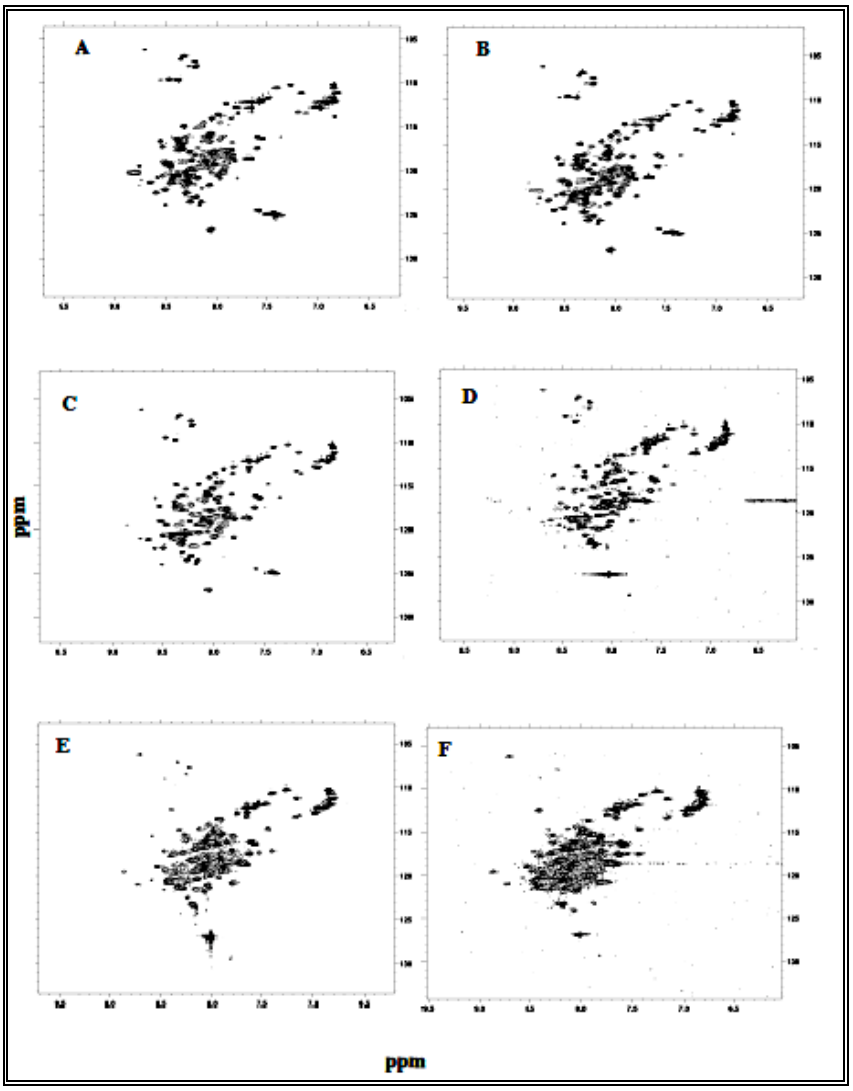


Figure 12: Plots of the two dimensional [^{15}N,^{1}H]-HSQC spectra of Y4-N-TM4-TM5. Spectra were recorded using 0.2 mM samples of the protein in 40mM phosphate buffer containing 40mM DTT in 5% LPPG at different pH such as (A) pH 5.0; (B) pH5.5; (C) pH 6.0; (D) pH 6.5; (E) pH 7.0 and (F) pH 7.5

7.2.8 Construction of larger transmembrane fragments of Y4 receptor

The double transmembrane fragments of Y4 receptor gave an moderate expression level in *E.coli*. We wanted to extend this methodology to express even larger fragments of Y4 –receptor. The first 3TM region of Y4-receptor (N-TM1-TM2-TM3) was cloned into pLC01 vector without TrpΔLE fusion to check whether direct expression as in the case of N-TM1-TM2 is possible. Unfortunately this fragment didn't give any

overexpression. Simultaneously another 3TM construct (TrpΔLE-TM4-TM5-TM6), and 4TM construct (TrpΔLE-TM4-TM5-TM6-TM7-C) were cloned and both the constructs did not give any expression. In order to achieve good expression of these larger fragments currently we were trying other fusion partners like streptococcal GB1 domain, MBP, NusA tag that can be cleaved from the fusion protein by 3C protease.

7.3 Conclusions

Here we report on the biosynthetic production of double TM constructs from the human Y4 receptor in *E.coli*. After systematic screening of growth conditions, host strains and cleavage conditions a yield of 1-2 mg of target protein can be achieved from 1 litre of culture. Because the amounts of double TM fragments were limited, the screening trials varied only the detergents while keeping other variables like protein, detergent concentration, temperature, pH and salt concentration constant. However for the N-TM4-TM5 fusion protein in LPPG micelle which gave better spectra among all the detergents which are screened, we have done a pH screen to check whether it helps to get well resolved spectra. Though the spectral quality varies with different pH range it could not lead to good quality of spectra. For the isolated 2TM proteins once when we get reasonable spectra with some detergent or detergent mixtures, fine tuning can be done with varying salt concentration, pH and detergent concentration.

While well-resolved ([^{15}N,^{1}H]-HSQC) spectra do not guarantee that a three dimensional structure can be determined, they may serve as highly sensitive monitors of purification and the entire sample preparation process and reflect the sample aggregation state. The 1.5 ppm dispersion of the amide proton chemical shifts is typical of native helical membrane proteins in micelles. Best chances for successful studies exists if more than 90% of all resonances are observed in a sufficient-quality ([^{15}N,^{1}H]-HSQC) spectrum. When this is not the case the amount of information available in later experiments will be severely compromised, limiting the possibility for successful assignments and structure determination. We are now in the process of fine tuning the detergent conditions to obtain a well dispersed ([^{15}N,^{1}H]-HSQC) spectra for the Y4 TM4-TM5 fragment so as to proceed with triple resonance experiments to obtain the backbone

assignments. In case of the Y4 TM6-TM7-C construct CHAPS and LDAO are found to be the more suitable detergent for the C3 protease cleavage. However, optimization is still required for buffer ingredients. Once cleavage conditions are optimized we will proceed with the sample optimization for NMR measurements and hopefully proceed with assignments and structure calculations.

The unique challenge in the study of membrane proteins lies in their inherent hydrophobicity and the difficulty to concentrate to levels high enough for most structural studies without aggregation. Sample optimization is demanding due to low yields in membrane protein production and the requirement of stable-isotope labeling. When the hydrophobic part of a transmembrane protein is too large or too small to match the hydrophobic bilayer thickness, the mismatch between the hydrophobic surface and the micelle dimensions may result in a perturbation of the protein structure to avoid hydration of the hydrophobic α-helices. The structural heterogeneity and potential exchange between the two populations of proteins provide a possible explanation for the observed NMR line broadening. Matching of the micelle dimensions to the protein's hydrophobic surface avoids exchange processes that reduce the completeness of the NMR observations. Mixing detergents at different ratios might be a way to systematically change the size and shape of detergent micelles, which leads to the rational design of mixed micelles. The relationship between micelle size and thickness and membrane protein conformational homogeneity, plays an important role for NMR structure determination of transmembrane proteins. However, a correlation between the physical properties of the detergent micelle and the likelihood of obtaining a membrane protein structure is currently not well understood. Hence an extensive detergent screening has to be done to obtain well resolved spectra suitable for the structural studies.

A major problem with the 2 TM approach is that the biological significance of the structures is difficult to prove. In order to take care of this problem we plan to add missing parts of the TM region as separate proteins to the sample. For some cases it could be shown to result in split receptors capable of signal transduction. For those split receptors assays to monitor their biological activity are available. The remaining part will

be added as a non-labelled protein, so that the 2-TM part under study is only visible in the NMR spectra. We are likely to observe changes in chemical shifts of 2 TM parts in presence and absence of the remainder of the receptor, but expect that the assignments can be adjusted. In addition to the structure of the 2TM segment in a biologically relevant context these experiments will also deliver information on the stability and assembly of the full-length receptor.

7.4 References

1. Balasubramaniam, A., *Clinical potentials of neuropeptide Y family of hormones.* Am J Surg, 2002. **183**(4): p. 430-4.
2. Weiss, H.M. and R. Grisshammer, *Purification and characterization of the human adenosine A(2a) receptor functionally expressed in Escherichia coli.* Eur J Biochem, 2002. **269**(1): p. 82-92.
3. White, J.F., et al., *Automated large-scale purification of a G protein-coupled receptor for neurotensin.* FEBS Lett, 2004. **564**(3): p. 289-93.
4. Krepkiy, D., et al., *Bacterial expression of functional, biotinylated peripheral cannabinoid receptor CB2.* Protein Expr Purif, 2006. **49**(1): p. 60-70.
5. Baneres, J.L., et al., *Structure-based analysis of GPCR function: conformational adaptation of both agonist and receptor upon leukotriene B4 binding to recombinant BLT1.* J Mol Biol, 2003. **329**(4): p. 801-14.
6. LaVallie, E.R. and J.M. McCoy, *Gene fusion expression systems in Escherichia coli.* Curr Opin Biotechnol, 1995. **6**(5): p. 501-6.
7. Cohen, L.S., et al., *Expression and biophysical analysis of two double-transmembrane domain-containing fragments from a yeast G protein-coupled receptor.* Biopolymers, 2008. **90**(2): p. 117-30.
8. Popot, J.L. and D.M. Engelman, *Helical membrane protein folding, stability, and evolution.* Annu Rev Biochem, 2000. **69**: p. 881-922.
9. Estephan, R., et al., *Biosynthesis and NMR analysis of a 73-residue domain of a Saccharomyces cerevisiae G protein-coupled receptor.* Biochemistry, 2005. **44**(35): p. 11795-810.
10. Ulfers, A.L., et al., *Structure of the third intracellular loop of the human cannabinoid 1 receptor.* Biochemistry, 2002. **41**(38): p. 11344-50.
11. Ulfers, A.L., A. Piserchio, and D.F. Mierke, *Extracellular domains of the neurokinin-1 receptor: structural characterization and interactions with substance P.* Biopolymers, 2002. **66**(5): p. 339-49.
12. Katragadda, M., J.L. Alderfer, and P.L. Yeagle, *Assembly of a polytopic membrane protein structure from the solution structures of overlapping peptide fragments of bacteriorhodopsin.* Biophys J, 2001. **81**(2): p. 1029-36.
13. Hjelmeland, L.M. and A. Chrambach, *Solubilization of functional membrane proteins.* Methods Enzymol, 1984. **104**: p. 305-18.
14. van Renswoude, J. and C. Kempf, *Purification of integral membrane proteins.* Methods Enzymol, 1984. **104**: p. 329-39.

15. le Maire, M., P. Champeil, and J.V. Moller, *Interaction of membrane proteins and lipids with solubilizing detergents.* Biochim Biophys Acta, 2000. **1508**(1-2): p. 86-111.
16. Konopka, J.B., S.M. Margarit, and P. Dube, *Mutation of Pro-258 in transmembrane domain 6 constitutively activates the G protein-coupled alpha-factor receptor.* Proc Natl Acad Sci U S A, 1996. **93**(13): p. 6764-9.
17. Dube, P. and J.B. Konopka, *Identification of a polar region in transmembrane domain 6 that regulates the function of the G protein-coupled alpha-factor receptor.* Mol Cell Biol, 1998. **18**(12): p. 7205-15.
18. Staley, J.P. and P.S. Kim, *Formation of a native-like subdomain in a partially folded intermediate of bovine pancreatic trypsin inhibitor.* Protein Sci, 1994. **3**(10): p. 1822-32.
19. Miroux, B. and J.E. Walker, *Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels.* J Mol Biol, 1996. **260**(3): p. 289-98.
20. Studier, F.W. and B.A. Moffatt, *Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes.* J Mol Biol, 1986. **189**(1): p. 113-30.
21. Guzman, L.M., et al., *Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter.* J Bacteriol, 1995. **177**(14): p. 4121-30.
22. Zou, C., F. Naider, and O. Zerbe, *Biosynthesis and NMR-studies of a double transmembrane domain from the Y4 receptor, a human GPCR.* J Biomol NMR, 2008. **42**(4): p. 257-69.
23. Zou, C., et al., *Studies of the structure of the N-terminal domain from the Y4 receptor - a G protein-coupled receptor - and its interaction with hormones from the NPY family.* Chembiochem, 2008. **9**(14): p. 2276-84.
24. Tian, C., et al., *Solution NMR spectroscopy of the human vasopressin V2 receptor, a G protein-coupled receptor.* J Am Chem Soc, 2005. **127**(22): p. 8010-1.
25. Cordingley, M.G., et al., *Cleavage of small peptides in vitro by human rhinovirus 14 3C protease expressed in Escherichia coli.* J Virol, 1989. **63**(12): p. 5037-45.
26. Cordingley, M.G., et al., *Substrate requirements of human rhinovirus 3C protease for peptide cleavage in vitro.* J Biol Chem, 1990. **265**(16): p. 9062-5.
27. Walker, P.A., et al., *Efficient and rapid affinity purification of proteins using recombinant fusion proteases.* Biotechnology (N Y), 1994. **12**(6): p. 601-5.

CURRICULUM – VITAE

K. Sowmini

Ph.D graduate
c/o Prof. Oliver Zerbe
Organic Chemistry Institute
University of Zürich
CH-8057 Zürich
E-Mail: sowminik@gmail.com

Telephone : +41 44 6353940
Mobile : +41 786044133

PROFILE

I am a Ph.D graduate with career ambition in the field of structural biology and immunology. I have completed my Ph.D in the field of Bio-NMR at the Organic Chemistry Institute, University of Zurich in January 2010.

EDUCATIONAL QUALIFICATIONS

YEAR	QUALIFICATION
Apr 2005 – Jan 2010	Worked as a research scholar in the field of Bio-NMR in the Institute of Organic Chemistry, University of Zurich, under the guidance of Prof. Oliver Zerbe**.** **https://www.unizh.ch/oci/group.pages/zerbe/index.html**
Aug 2003- Mar 2005	Worked as a project assistant with Prof. Raghavan Varadarjan (Molecular Biophysics Unit, Indian Institute of Science, Bangalore) in a HIV vaccine development project funded by National Institute of Health (NIH). **http://mbu.iisc.ernet.in/~rvgrp**
2000 – 2002	Completed my Master's Degree in Biochemistry from Dr. ALM PG Institute of Basic Medical Sciences (University of Madras) with an aggregate of 72.5%.
1996 – 1999	Completed my Bachelor's Degree in Chemistry from Anna Adarsh college for women, University of Madras with an aggregate of 74.78 %.
1994 – 1996	Completed my Diploma in Medical Laboratory Technology (DMLT) from King Institute of Preventive Medicine.

RESEARCH EXPERIENCE

Apr 2005 - Dec 2009 Ph.D in University of Zurich / Switzerland **Advisor: Prof. Oliver Zerbe**

I was involved in two projects, the first project focused on the interactions of bacterial lipopolysaccharide with different types of polymyxins studied by NMR spectroscopy and the second project involved recombinant production of domains from neuropeptide Y4 receptor, a member of G-protein coupled receptors (GPCRs), and their structural and functional exploration by multi-dimensional NMR techniques.

Aug 2003 - Mar 2005 Project Assistant in IISc Bangalore / India Advisor: Prof. Raghavan Varadarajan.

The project focused on developing stable complexes of glycoprotein 120 of HIV and CD4 of T-cells in order to develop HIV antibodies. I was involved in cloning, bacterial and mammalian expression of several mutants and shorter constructs of CD4, gp120-CD4, trimeric gp140 constructs that could lead to the formation of stable gp120 complex with its epitopes exposed.

Dec 2001 - Jun 2002 Masters research project University of Madras / India Advisor: Prof. R. Selvam and Prof. P. Varalakshmi.

We carried out a study to check whether the oxalate rich soybeans lead to kidney stone formation in animal models.

ACADEMIC ACHIEVEMENTS

GATE - 2002 Life science discipline

Qualified in the Graduate Aptitude Test in Engineering held in February - 2002 in the Life science discipline. Percentile Score - 93.03.

ICMR examination - 2003

Qualified the ICMR /JRF- All India entrance examination conducted by the Indian council of medical research held in July 2003.

University of Madras competitive entrance test

Selected for admission to M.Sc. Biochemistry (two years) through all India entrance exam conducted by University of Madras, Tamilnadu, India.

SKILLS AND TECHNICAL EXPERTISE

Molecular biology and Cellular biology

PCR, site-directed mutagenesis, recombinant protein purification by affinity, ion-exchange, size-exclusion and reverse phase chromatography, western blotting, ELISA, slot blotting, ESI and MALDI-TOF mass spectrometry, mammalian cell culture techniques (cryopreservation, transfection and maintenance of cell lines).

Biochemistry, Microbiology and Pathology

Basic biochemistry assays, isolation and characterization of microbial strains, gram staining, AFB staining, culture and sensitivity in blood and urine, identification of parasites, fungal cultures, routine clinical hematological tests, urinary stone analysis, routine clinical laboratory tests for pathology.

Structural biology

NMR data acquisition and processing (BRUKER), spectra analysis (XEASY, CARA, Sparky), structure calculation and refinement (CYANA), isotopic labeling of proteins with ^{2}H, ^{13}C and ^{15}N.

PUBLICATIONS

1. Jiri Mares, **Sowmini Kumaran**, Marina Gobbo, Zerbe Oliver*. "Interactions of lipopolysaccharide and polymyxin studied by NMR spectroscopy" J Biol Chem, *2009* **284**, 11498-506.

2. Zou Chao, **Sowmini Kumaran**, Reto Walser, Zerbe Oliver*. "Properties of N-terminal domains from Y receptors probed by NMR spectroscopy". J Pept Sci. 2009 Mar;15(3):184-91.

3. Zou Chao, **Sowmini Kumaran**, Stefan Markovic, Reto Walser, Zerbe Oliver*. "Studies of the structure of the N-terminal domain from Y4 receptor, a G-protein coupled receptor, and its interaction with hormones from NPY family". ChemBioChem. 2008 Sep 22;9(14):2276-84

4.Deepak Sharma, M.M. Balamurali, Kausik Chakraborty, **Sowmini Kumaran**, Sadasivam Jeganathan, Umar Rashid, Paolo Ingallinella, Raghavan Varadarajan*."Protein minimization of the gp120 binding region of human CD4" Biochemistry 44:16192-16202 (2005)

POSTERS, TRAININGS AND AWARDS

- "Interactions of polymyxins and lipopolysaccharides studied by NMR spectroscopy ".
  Sowmini Kumaran, Jiri Mares, Oliver Zerbe - Swiss Chemical Society Fall Meeting 2008 held in University of Zurich, Switzerland on September 11, 2008.

- Received a travel award of the Swiss Academy of Sciences (SCNAT) Chemistry Travel Award in 2009 to meet the traveling expenses for participation in international conferences.

- Attended and gave my work presentation in the "Protein science summer school 2009" conducted by the University of Copenhagen, Department of Biology in Spetses, Greece from May 28th - June 5th, 2009.

WORK EXPERIENCE

- **2006 to 2009 -** NMR service measurement for organic chemistry institute, University of Zurich.
- **Dec 2002 to July 2003 -** Worked as a biochemist in P.S. Medicare and Diagnostic Center, Ashok Nagar, Chennai-83.
- **Jan 2000 - Aug 2000** - Worked as a laboratory technician in Govt. Women and children hospital Egmore, Chennai- 8 in a (NACO) National AIDS control organization project
- **Aug 1997 - June 2002 -** Worked as a part time laboratory technician in Akkurate diagnostic centre, Annanagar, Chennai.

PERSONAL DETAILS

Date of Birth: 13th May, 1976
Sex: Female
Marital status: Single
Nationality: Indian
Languages Known: English, Tamil, and Malayalam.
House Address: Oerlikoner strasse-44/26
CH– 8057, Zurich, Switzerland

REFERENCES

Prof. Dr. Oliver Zerbe	**Prof. Dr. Raghavan varadarajan**	**Prof. Dr. P. Sachidhanandham**
Organic Chemistry Institute University of Zurich Winterthurerstrasse 190 CH-8057 Zurich Switzerland	Molecular Bio physics Unit, Indian institute of science Bangalore - 560 012 Karnataka India	University of Madras Department of Medical Biochemistry Taramani campus Chennai,India
Phone: 0041 44 6354263	Phone: 0091-80-2293 2612	Phone: 0091-44-24900025
Email: oliver.zerbe@oci.uzh.ch	Email: varadar@mbu.iisc.ernet.in	Email: psachdanandam2000@yahoo.co.in

Acknowledgements

In the last part of my thesis I wish to express my sincere thanks to the people who helped me to make this thesis done. First and foremost I would like to express my deep and sincere gratitude to my supervisor Professor. Dr. Oliver Zerbe, for giving me the opportunity to carry out this thesis in his laboratory and his commitment and excellent teaching of the complicated matter of NMR spectroscopy in an understandable manner and encouragement, guidance, and support from the initial to the final level. Thank you very much Oliver…

I am really blessed to have such wonderful colleagues. Our group has been a source of friendship as well as good advice and collaboration. Especially I would like to thank wonderful trios of the group. Reto is a very great friend as well as a good mentor. He is always there to talk about my project and ask me good questions, which helped me to think through my problems and indeed patiently listened my boring stories and horrible songs. He is a friend whom I can trust that he will come immediately to help when needed. It would be really unfair if I didn't thank him for proof reading my thesis chapters and german translation of my thesis summary.

Then my great thanks goes to another wonderful lovely friend Chao for his warm friendship, for his help to solve experimental problems alongside having fruitful discussions. He taught me how to work hard and gave insightful comments on my project and provided moral support during difficult times. Chao, you are simply great! I also take this opportunity to thank Jiri for his nice friendship, countless timely help and especially for tolerating my quarrels with him. Many thanks goes to my other wonderful labmates Alexey, Sivakami, Stefan, Harsha, Christina, Svava, Xuan, Tanja, Nadine, Randall, Jacapo, Martin, for the nice atmosphere and fun in the laboratory.

I also appreciate the kind support that our technicians Simon Jurt and Nadja Bross have offered during the NMR service work. I am grateful to Professor Dr. John Robinson and his group members for allowing us to use their equipments and chemicals and Dr. Katja Zerbe for providing many plasmids and bacterial strains and for her nice advises on several molecular biology related problems. A great thanks also goes to Professor. Dr. Nathan Luedtke for correcting my thesis. Of course special thanks should also be given to mass spectroscopy service and administrative office team.

My time at Zurich was made enjoyable in large part due to my Indian friends who became a part of my life. Isai, Naresh, Kanakaraj, Srini, Ajit, Kalpana, Pradeep, Smitha, Jimy, Saju, Sarath, Anshu, Gopi, Amit, Ram, Yeswanth, Harsha, Payal, Anu, Shreya, Chitra, Rajkumar, Sathya, Dinesh, Kavitha, Bala, Illavarasi. They have showered love, care, affection etc and made to realize that friendship isn't a one big thing - it's a million little things. I enjoyed the entire outings, birthday parties with egg masalas, numerous funny games and delicious dinner times with you all. Thanks a lot dear friends. Moments shared with you guys will always be cherished. I would also like to thank Prasanna`s mother and Kalpana`s father for their kindness and delicious food during the stressful period of thesis writing. In this time, I should definitely mention about two girls Jocia and Prasanna who played a major role in tuning my life in many aspects. We have been friends together in sunshine and in shade. Sometimes you put walls up not to keep people out, but to see who cares enough to break them down. These two friends are of that type. They are truly great friends hard to find, difficult to leave, and impossible to forget.

I am indebted to my lovely mother, who gave me life, raised me, taught me and loves me no matter what happens. God could not be everywhere, so he created mothers. This is a real truth. All that I am or ever hope to be, I owe to my beautiful angel mother. I would like to thank my lovely father for showering countless blessings from heaven. Really I miss you a lot Accha.. A very big hug to my brother Mohan for his immense efforts to fulfill all my needs, for his endless patience and encouragement when it was most required. Anna, I am smiling because you are my brother. I am laughing because there is nothing you can do about it!. If I am given the chance to select the best brother in the world, It would be you only! I also want to thank my sister in law, her sweet kids Harish and Thusara, all my relatives, and my friends for their immense love and support. I would like to thank all my previous teachers for their valuable teachings and blessings. Finally I would like to thank God Almighty. “Faith is not knowing God can…it’s knowing that he will.” Thanks a lot Ganesha…